PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-324077
(43)Date of publication of application : 08.11.2002

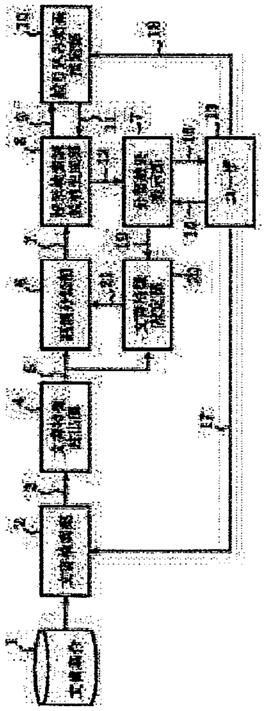
(51)Int.Cl. G06F 17/30

(21)Application number : 2001-126541
(22)Date of filing : 24.04.2001

(71)Applicant : MITSUBISHI ELECTRIC CORP
(72)Inventor : NAGAI AKITO
TAKAYAMA YASUHIRO
SUZUKI KATSUSHI

(54) APPARATUS AND METHOD FOR DOCUMENT RETRIEVAL

(57)Abstract:
PROBLEM TO BE SOLVED: To solve the problem that it is difficult for a conventional apparatus for document retrieval to select a searching word to efficiently narrow a search.
SOLUTION: An apparatus for document retrieval comprises a document retrieval part 2 to search documents, a document feature extracting part 4 to output a document vector group 5, a topic classifying part 6 to prepare a topic by classifying the group 5, a narrowing effect presumption part 10 to calculate narrowing effect indicator 11,a generating part 8 for choice of retrieving word for provision to select the retrieval word with high indicator 11 and output it as a candidate of retrieval word 12 for provision, a classifying result providing part 14 to provide the candidate 12 and indicator 11 topic-by topic and a document feature setting part 20 to change the group 5.



[Claim(s)]

[Claim 1]A document retrieval means which searches and outputs a document which suits a search condition from a document set, A document feature extraction means which outputs a document vector set which computes statistical dignity of said word and is obtained from said dignity based on the frequency of occurrence of a word described by said document, A subject sorting means which outputs document information which creates subject and belongs to said subject by classifying said document vector set according to similarity between document vectors, and a search term with importance of said subject, The narrowing-down effect estimation means which computes and outputs a narrowing effect indicator of a search term which belongs to said subject with reference to said document information and said search term with importance, A presentation search term candidate creating means which gives said narrowing effect indicator to said search term, said narrowing effect indicator chooses said high search term, considers it as a presentation search term candidate, and outputs document information corresponding to this presentation search term candidate and this presentation search term candidate, A classification result presenting means urged to match said presentation search term candidate and said subject, to show with said narrowing effect indicator, and to input either directions information or selection information and both, A document retrieval system provided with a document feature setting means which changes and outputs a document vector included in said document vector set based on said selection information.

[Claim 2]The document retrieval system according to claim 1, wherein the narrowing-down effect estimation means computes and outputs a narrowing effect indicator to one or more subjects at least.

[Claim 3]The document retrieval system according to claim 1, wherein a classification result presenting means presents a presentation search term candidate belonging to subject and this subject in the form of a procession and shows each element of said procession a narrowing effect indicator.

[Claim 4]Extract a word by which this document is characterized as a subject classification item from a document which a document retrieval means outputs, and a text relevant to this subject classification item is referred to, Learn a weight vector to said subject classification item, have a subject classification item acquisition means which outputs said subject classification item and said weight vector to a subject sorting means, and said subject sorting means, The document retrieval system according to claim 1 creating subject by classifying a document vector set based on said subject classification item and said weight vector.

[Claim 5]The document retrieval system according to claim 4, wherein a subject classification item acquisition means extracts a subject classification item from a document which a document retrieval means outputs based on the frequency of occurrence of a word described by this document.

[Claim 6]The document retrieval system according to claim 4, wherein a subject classification item acquisition means extracts a subject classification item from a document which a document retrieval means outputs with reference to a tag described by this document.

[Claim 7]Compute a document vector from a document specified via a classification result presenting means, have a specified document feature extraction means outputted to a document feature setting means, and said document feature setting means, The document retrieval system according to claim 1 changing said document vector set based on said document vector which said specified document feature extraction means outputs, and a document vector set which a document feature extraction means outputs.

[Claim 8]The 1st recording device that defines a word relevant to a predetermined word and is recorded as a related term, Have a related term setting-out means to extract said related term corresponding to a specified search term from said 1st recording device, and to output to a document feature setting means, and said document feature setting means, The document retrieval system according to claim 1 characterized by changing a document vector set based on selection information inputted from said related term and a classification result presenting means.

[Claim 9]The 2nd recording device that records creation knowledge of a retrieval request statement, and this 2nd recording device are referred to, Create a retrieval request statement corresponding to a search term which a presentation search term candidate creating means outputted, have a retrieval-required preparing means outputted to a document retrieval means, and said presentation search term candidate creating means, The document retrieval system according to claim 1 choosing said search term outputted to said retrieval-required preparing means based on a narrowing effect indicator.

[Claim 10]The document retrieval system according to claim 9, wherein a presentation search term candidate creating means chooses two or more search terms, and outputs them to a retrieval-required preparing means and this retrieval-required preparing means creates a retrieval request statement from a logical operation to said two or more search terms.

[Claim 11]A document-retrieval step which searches and outputs a document which suits a search condition from a document set, A document feature extraction step which outputs a document vector set which computes statistical dignity of said word and is obtained from said dignity based on the frequency of occurrence of a word described by said document, A subject classification step which outputs document information which creates subject and belongs to said subject by classifying said document vector set according to similarity between document vectors, and a search term with importance of said subject, The narrowing-down effect estimating step which computes and outputs a narrowing effect indicator of a search term which belongs to said subject with reference to said document information and said search term with importance, A presentation search term candidate generation step which gives said narrowing effect indicator to said search term, said narrowing effect indicator chooses said high search term, considers it as a presentation search term candidate, and outputs document information corresponding to this presentation search term candidate and this presentation search term candidate, Said presentation search term candidate and said subject are matched, and it shows with said narrowing effect indicator, and is based on a classification result presentation step urged to input either directions information or selection information and both and said selection information, A document retrieval method which has the document feature setting step which changes and outputs a document vector included in said document vector set.

[Claim 12]The document retrieval method according to claim 11, wherein the narrowing-down effect estimating step computes and outputs a narrowing effect indicator to one or more subjects at least.

[Claim 13]The document retrieval method according to claim 11, wherein a classification result presentation step presents a presentation search term candidate belonging to subject and this subject in the form of a procession and shows each element of said procession a narrowing effect indicator.

[Claim 14]Extract a word by which this document is characterized as a subject classification item from a document which a document-retrieval step outputs, and a text relevant to this subject classification item is referred to, Learn a weight vector to said subject classification item, have a subject classification item acquisition step which outputs said subject classification item and said weight vector, and a subject classification step, The document retrieval method according to claim 11 creating subject by classifying a document vector set based on said subject classification item and said weight vector.

[Claim 15]The document retrieval method according to claim 14, wherein a subject classification item acquisition step extracts a subject classification item from a document which a document-retrieval step outputs based on the frequency of occurrence of a word described by this document.

[Claim 16]The document retrieval method according to claim 14, wherein a subject classification item acquisition step extracts a subject classification item from a document which a document-retrieval step outputs with reference to a tag described by this document.

[Claim 17]Have a specified document feature extraction step which computes and outputs a document vector from a document specified via a classification result presentation step, and the document feature setting step, The document retrieval method according to claim 11 changing said document vector set based on said document vector which said specified document feature extraction step outputs, and a document vector set which a document feature extraction step outputs.

[Claim 18]The 1st record step that defines a word relevant to a predetermined word and is recorded as a related term, The document retrieval method according to claim 11, wherein it has a related term setting step which extracts and outputs said related term corresponding to a specified search term and the document feature setting step changes a document vector set based on selection information inputted from said related term and a classification result presentation step.

[Claim 19]It has the 2nd record step that records creation knowledge of a retrieval request statement, and a retrieval-required creation step which creates and outputs a retrieval request statement corresponding to a search term which a presentation search term candidate generation step outputted, The document retrieval method according to claim 11, wherein said presentation search term candidate generation step chooses said search term outputted based on a narrowing effect indicator.

[Claim 20]The document retrieval method according to claim 19, wherein a presentation search term candidate generation step chooses and outputs two or more search terms and a retrieval-required creation step creates a retrieval request statement from a logical operation to said two or more search terms.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Detailed Description of the Invention]

[0001]

[Field of the Invention]This invention relates to the document retrieval system and document retrieval method which can perform a search of a document efficiently by [which is a bigger unit than a document] narrowing down for every subject, showing an effect and making selection of an additional search term easy.

[0002]

[Description of the Prior Art]When the electronic document etc. which are recorded and managed are searched to the HTML document which can be perused using the Internet, or a large-scale text database, The amount of information of the document information acquired as search results increases dramatically, and many time and labors are needed for discovery of the document for which a user asks. For this reason, according to the contents, classify the document information acquired as search results so that a user can search the target document efficiently, or. The demand to the art which supports discovery of the document for which a user who shows a user the search term candidate who adds for narrowing retrieval asks is increasing.

[0003]About the art in which a user can search the target document efficiently, JP,H9-231238,A "text browsing result display method and device." There are (document 1 being called hereafter) and art indicated by JP,H11-213000,A "storage which stored the interactive information retrieval method, the device, and the interactive information retrieval program" (document 2 is called hereafter).

[0004]The art indicated in document 1 carries out the theme classification of the search results by fuzzy clustering, and shows a user the category by which the theme classification was carried out with the group of a search term. The art indicated in document 2 classifies search results according to clustering, shows a user the classified category with the group of a search term, classifies the category specified by a user into a subcategory further, and enables interactive search.

[0005]being related with the art which supports narrowing retrieval -- "development of the interface for search-results narrowing down in a WWW search service" (Information Processing Society of Japan and a human interface study group (HI76-5).) of Hayakawa and others In pp.25-1998 and the following, there is art which set to call document 3 and was indicated.

[0006]The art indicated in document 3 provides the interface which visualizes the information how many search-results numbers a search term can narrow down, and is shown to a user, in order to be able to grasp easily the relation between a search term and a search-results document set. The narrowing-down result of a search term is visualized using a document word matrix, The matrix which has a word and document in a row and column, respectively is shown to a direct user in the form of a table, the dignity of the word to document is expressed with the luminosity of the cell of a matrix, and it enables it to look through the narrowing-down effect of an additional search term.

[0007]There is art indicated by JP,H11-85764,A "storage which stored the statistical estimation program of the statistical estimation method of the search-results number, the device, and the search-results number" (document 4 is called hereafter) about the conventional technology described in document 3. The art indicated in document 4 is indicated about the statistical estimation method of the search-results number for knowing how the number of search results will change, when the search term for narrowing down search results further in document 3 is added.

[0008]Drawing 10 is a lineblock diagram showing the conventional document retrieval system. It is a lineblock diagram of the estimating device which is an example of the operation indicated by document 4.

An estimating device for the document retrieval according [on drawing 10 and / 101] to a full-text search, A search term for 102 to search document, the text database with which document is recorded and managed 103, The search-results total of document in which 104 is outputted from the text database 103, The document sample set which 105 extracts 50 affairs at random out of search results, and is outputted from the text database 103, The document-word-matrix generation part in which the document sample set 105 is inputted into and 106 counts the appearance frequency of each word about each document, and 107 are the document word matrices which the document-word-matrix generation part 106 generated. The document sample set 105 is not limited to 50 affairs, and can be suitably set up according to a situation.

[0009]The search term candidate presentation part in which 108 calculates word significance to the document word matrix 107 in drawing 10, The search term candidate to whom 109 is outputted from the search term candidate presentation part 108, the search term selecting part by which 110 outputs the search term candidate 109 to a monitor etc., The selection signal into which 111 is inputted when a user chooses the search term candidate 109, The additional search term to which 112 is outputted from the search term selecting part 110 based on the selection signal 111, The number-of-cases estimating part in which 113 calculates the incidence of the additional search term 112 based on the document word matrix 107 and the additional search term 112, and 114 are the presumed numbers produced by being outputted from the number-of-cases estimating part 113, and multiplying the search-results total 104 by the incidence. Drawing 11 is an explanatory view showing an example of the document word matrix 107 in the conventional document retrieval system.

[0010]Next, operation is explained. If the search term 102 inputs into the text database 103, based on the search term 102, document recorded and managed will be searched to the text database 103. The text database 103 extracts 50 affairs at random out of search results, and outputs them to the document-word-matrix generation part 106 as the document sample set 105, and it outputs the search-results total 104 to the number-of-cases estimating part 113. By counting the number of times to which each word appears about each document based on the document sample set 105, the document-word-matrix generation part 106 generates the document word matrix 107 shown in drawing 11, and outputs it to the search term candidate presentation part 108 and the number-of-cases estimating part 113. In the document word matrix 107, a row and column shows a literature identifier and a search term list, respectively, and the value of the cell of a table shows the number of times to which the search term of a sequence [be / it / under / of corresponding document of a line / correspondence] appears.

[0011]The search term candidate presentation part 108 calculates the word significance which shows how important words arbitrary about arbitrary document are except for words, such as a particle and a pronoun, from the document word matrix 107. Word significance is an index which becomes high and becomes low with the word which has appeared by many document conversely with the word which has appeared in specific document intensively. Based on the calculated word significance, word significance outputs the search term candidate presentation part 108 to the search term selecting part 110 by making a high word into the search term candidate 109.

[0012]The search term selecting part 110 outputs the search term candidate 109 to a monitor etc., makes a user choose arbitrary search terms from the search term candidate 109, and is made to input as the selection signal 111. The search term selecting part 110 is outputted to the number-of-cases estimating part 113 as the additional search term 112 based on the selection signal 111 which the user inputted.

[0013]The search-results total 104, the document word matrix 107, and the additional search term 112 input into the number-of-cases estimating part 113. The number-of-cases estimating part 113 counts the number of the lines whose sequences of the additional search term 112 of the document word matrix 107 are not "0", and the incidence of the additional search term 112 is obtained by **(ing) this with the total number of lines. The number-of-cases estimating part 113 outputs the presumed number 114 produced by multiplying the incidence of the additional search term 112 by the search-results total 104.

[0014]

[Problem(s) to be Solved by the Invention]Since the conventional document retrieval system is constituted as mentioned above, about the narrowing-down effect of a search term, according to the conventional technology indicated by document 1 and document 2, classified the document of search results, and have shown the user the search term in the classified category by it, but. Since there was no information on the narrowing-down effect corresponding to the shown search term when choosing in order to use the shown search term for narrowing retrieval, SUBJECT that it was difficult to choose the search term for carrying out narrowing retrieval efficiently occurred.

[0015]Although the conventional document retrieval system has shown the user the narrowing-down effect of a search term per document by the conventional technology indicated by document 3 about visualization of search results, When search results became a scale which is about thousands, the display of the narrowing-down effect of the whole search results by a document word matrix became very difficult, and SUBJECT that list nature was missing occurred.

[0016]By the conventional technology indicated by document 4, the conventional document retrieval system about visualization of search results. From the incidence of the additional search term which outputs the document sample set from search results, and corresponds for every document based on the document sample set, since the narrowing-down number to the whole search results is presumed, Since it became the presentation for every additional search term when showing a user the presumed narrowing-down number, the display of the narrowing-down effect of the whole search results became very difficult, and SUBJECT that list nature was missing occurred.

[0017]Although the conventional document retrieval system is carrying out specific classification of the category specified in order to narrow down to the target document to the subcategory by the conventional technology indicated by document 2 about the narrowing retrieval to the target document, The target document does not exist altogether in the specified category, Since the document for which a user asks existed in other categories, the document obtained by narrowing retrieval was limited only to the document which exists in the specified category and the retrieval omission of the target document arose, SUBJECT that it was difficult to perform narrowing retrieval again occurred.

[0018]By the conventional technology indicated by document 3 and document 4, the conventional document retrieval system about the narrowing retrieval to the target document. Since AND retrieval which adds the search term to first-time search results is performed, the retrieval object of narrowing retrieval, It was limited to the document of search results to the last search term, search of a document for it to be scattered to whole sentence document space became impossible, and SUBJECT that it was difficult to perform narrowing retrieval again occurred.

[0019]As opposed to the search term which it was made in order that this invention might solve above SUBJECT, and is shown for narrowing retrieval, The index showing the narrowing-down effect is given, selection of an additional search term is made easy, and it aims at acquiring the document retrieval system and document retrieval method which make a user's narrowing retrieval efficient.

[0020]This invention classifies the document of search results, and makes subject the classified document set, and the narrowing-down effect of a search term is shown to a user for every subject which is a bigger unit than a document, It aims at acquiring the document retrieval system and document retrieval method which improve the list nature of the narrowing-down effect to the whole document of search results.

[0021]When this invention carries out reclassification of the first-time search results using the subject which the user specified as feedback information, or the information on a search term, As the document of the purpose distributed between subjects is brought together in one subject, it aims at acquiring the document retrieval system and document retrieval method which make narrowing down of search results efficient.

[0022]

[Means for Solving the Problem]A document retrieval means which a document retrieval system concerning this invention searches a document which suits a search condition from a document set, and is outputted, A document feature extraction means which outputs a document vector set which computes statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of a word described by document, A subject sorting means which outputs document information and a search term with importance of subject which create subject and belong to subject by classifying a document vector set according to similarity between document vectors, The narrowing-down effect estimation means which computes and outputs a narrowing effect indicator of a search term which belongs to subject with reference to document information and a search term with importance, A presentation search term candidate creating means which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, A classification result presenting means urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Based on selection information, it has a document feature setting means which changes and outputs a document vector included in a document vector set.

[0023]The narrowing-down effect estimation means computes a narrowing effect indicator, and it is made to output a document retrieval system concerning this invention to one or more subjects at least.

[0024]A document retrieval system concerning this invention presents a presentation search term candidate to whom a classification result presenting means belongs to subject and the subject concerned in the form of a procession, and shows each element of a procession a narrowing effect indicator.

[0025]A document retrieval system concerning this invention extracts a word by which the document concerned is characterized as a subject classification item from a document which a document retrieval means outputs, and a text relevant to a subject classification item is referred to, Learn a weight vector to a subject classification item, have a subject classification item acquisition means which outputs a subject classification item and a weight vector to a subject sorting means, and a subject sorting means, Subject is created by classifying a document vector set based on a subject classification item and a weight vector.

[0026]A document retrieval system concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition means is outputted from a document retrieval means based on the frequency of occurrence of a word described by the document concerned.

[0027]A document retrieval system concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition means is outputted from a document retrieval means with reference to a tag described by the document concerned.

[0028]A document retrieval system concerning this invention computes a document vector from a document specified via a classification result presenting means, It has a specified document feature extraction means outputted to a document feature setting means, and a document feature setting means changes a document vector set based on a document vector outputted from a specified document feature extraction means, and a document vector set outputted from a document feature extraction means.

[0029]The 1st recording device that a document retrieval system concerning this invention defines a word relevant to a predetermined word, and is recorded as a related term, It has a related term setting-out means to extract a related term corresponding to a specified search term from the 1st recording device, and to output to a document feature setting means, and a document feature setting means changes a document vector set based on selection information inputted from a related term and a classification result presenting means.

[0030]The 2nd recording device on which a document retrieval system concerning this invention records creation knowledge of a retrieval request statement, With reference to the 2nd recording device concerned, a retrieval request statement corresponding to a search term which a presentation search term candidate creating means outputted is created, It has a retrieval-required preparing means outputted to a document retrieval means, and a presentation search term candidate creating means chooses a search term outputted to a retrieval-required preparing means based on a narrowing effect indicator.

[0031]A presentation search term candidate creating means chooses two or more search terms, and outputs a document retrieval system concerning this invention to a retrieval-required preparing means, and a retrieval-required preparing means creates a retrieval request statement from a logical operation to two or more search terms.

[0032]A document-retrieval step which a document retrieval method concerning this invention searches a document which suits a search condition from a document set, and is outputted, A document feature extraction step which outputs a document vector set which computes statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of a word described by document, A subject classification step which outputs document information and a search term with importance of subject which create subject and belong to subject by classifying a document vector set according to similarity between document vectors, The narrowing-down effect estimating step which computes and outputs a narrowing effect indicator of a search term which belongs to subject with reference to document information and a search term with importance, A presentation search term candidate generation step which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, A classification result presentation step urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Based on selection information, it has the document feature setting step which changes and outputs a document vector included in a document vector set.

[0033]The narrowing-down effect estimating step computes a narrowing effect indicator, and it is made to output a document retrieval method concerning this invention to one or more subjects at least.

[0034]A document retrieval method concerning this invention presents a presentation search term candidate to whom a classification result presentation step belongs to subject and the subject concerned in the form of a procession, and shows each element of a procession a narrowing effect indicator.

[0035]A document retrieval method concerning this invention extracts a word by which the document concerned is characterized as a subject classification item from a document which a document-retrieval step outputs, and a text relevant to a subject classification item is referred to, A weight vector to a subject classification item is learned, it has a subject classification item acquisition step which outputs a subject classification item and a weight vector, and a subject classification step creates subject by classifying a document vector set based on a subject classification item and a weight vector.

[0036]A document retrieval method concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition step is outputted from a document-retrieval step based on the frequency of occurrence of a word described by the document concerned.

[0037]A document retrieval method concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition step is outputted from a document-retrieval step with reference to a tag described by the document concerned.

[0038]A document retrieval method concerning this invention has a specified document feature extraction step which computes and outputs a document vector from a document specified via a classification result presentation step, The document feature setting step changes a document vector set based on a document vector outputted from a specified document feature extraction step, and a document vector set outputted from a document feature extraction step.

[0039]The 1st record step that a document retrieval method concerning this invention defines a word relevant to a predetermined word, and is recorded as a related term, It has a related term setting step which extracts and outputs a related term corresponding to a specified search term, and a document vector set is changed based on selection information which the document feature setting step inputted from a related term and a classification result presentation step.

[0040]The 2nd record step on which a document retrieval method concerning this invention records creation knowledge of a retrieval request statement, It has a retrieval-required creation step which creates and outputs a retrieval request statement corresponding to a search term which a presentation search term candidate generation step outputted, and a presentation search term candidate generation step chooses a search term outputted based on a narrowing effect indicator.

[0041]A presentation search term candidate generation step chooses and outputs two or more search terms, and a document retrieval method concerning this invention creates a retrieval request statement from a logical operation [as opposed to two or more search terms in a retrieval-required creation step].

[0042]

[Embodiment of the Invention]Hereafter, one gestalt of implementation of this invention is explained.

Embodiment 1. drawing 1 is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 1. The HTML document which 1 is a document set used as a retrieval object in drawing 1, for example, can be perused using the Internet, Electronized texts, such as an electronic document recorded and managed, are consisted of by the E-mail which can be transmitted and received using the Internet, and the large-scale text database recorded on the recorder or the recording medium. The document retrieval part which searches a document from the document set 1 according to conditions predetermined in 2 (document retrieval means), The search-results document set which is the result of the document retrieval part 2 searching 3 from the document set 1 according to predetermined conditions, and 4 are document feature extraction parts (document feature extraction means) which create the document vector corresponding to the search-results document set 3 which the document retrieval part 2 outputted. The dignity of each word for every document is expressed as a document vector in the form of a vector.

[0043]The document vector set to which 5 is outputted in drawing 1 based on the document vector which the document feature extraction part 4 creates, The subject classification part which creates subject because 6 classifies the document vector set 5 into two or more sets according to the similarity between the document vectors computed based on the document vector set 5 which the document feature extraction part 4 outputted (subject sorting means), The search term set with importance outputted with the document information of each subject into which 7 was classified according to the subject classification part 6, and 8 are presentation search term candidate generation parts (presentation search term candidate creating means) which choose a presentation search term candidate from the search term set 7 with importance which the subject classification part 6 outputted on a predetermined standard. With the predetermined standard for choosing a presentation search term candidate, top 1 constant is chosen, for example in order of importance.

[0044]The search term set with importance of each subject to which 9 is outputted with the document information of each subject from the presentation search term candidate generation part 8 in drawing 1, The narrowing-down effect estimating part which presumes the narrowing-down effect of a search term that 10 belongs to the subject which the user specified from the search term set 9 with importance (the narrowing-down effect estimation means), The narrowing effect indicator which the narrowing-down effect estimating part 10 used as an index for 11 to presume the narrowing-down effect computed, The presentation search term candidate who outputs with the document information belonging to the subject which the presentation search term candidate generation part 8 chooses 12 based on the narrowing effect indicator 11, and corresponds, and the narrowing effect indicator 11, The user to whom 13 operates a document retrieval system, the classification result presentation part which 14 matches the presentation search term candidate 12 with each subject, and is shown to the user 13 with the narrowing effect indicator 11 (classification result presenting means), The classification result visualized in order that the classification result presentation part 14 might show the user 13 15, and 16 are directions information which the user 13 transmits to the classification result presentation part 14.

[0045]The search condition as which the user 13 inputs 17 into the document retrieval part 2 in drawing 1, The subject specified by the user 13 who the user 13 narrows down 18 and inputs into the effect estimating part 10, The selection information of the user 13 to whom 19 is outputted from the classification result presentation part 14 when the user 13 points to the reclassification of search results, The document feature set part (document feature setting means) by which 20 changes the dignity to a document vector or a search term based on the user's 13 selection information 19, and 21 are changed document vectors which the document feature set part 20 outputs to the subject classification part 6.

[0046]Drawing 2 is a flow chart explaining operation of the document retrieval system by this embodiment of the invention 1. Drawing 3 is an explanatory view showing an example of the document vector in this embodiment of the invention 1. Drawing 4 is an explanatory view showing an example of the search term-subject conversion table in this embodiment of the invention 1.

[0047]Next, operation is explained. First, in step ST1, the user 13 inputs the search condition 17 into the document retrieval part 2. The search condition 17 is a logical formula of a search term or two or more search terms, for example. Next, in step ST2, the document retrieval part 2 searches the document of the document set 1 based on the inputted search condition 17, and outputs the search-results document set 3 which suits the search condition 17 (document-retrieval step). The HTML document which can peruse the document set 1 used as a retrieval object, for example using the Internet, It is the text in which the electronic document etc. which are recorded and managed were electronized by the E-mail which can be transmitted and received using the Internet, and the large-scale text database recorded on the recorder or the recording medium. If search by the search condition 17 is possible for the document retrieval part 2, it is good, for example, the full-text search engine etc. which are generally used in the Internet may be used for it. The document retrieval part 2 is outputted with the search-results document set 3 by making information, including the place of the document ID number for specifying various information about the document of search results, for example, a document, or a document file, the title of a document, etc., into document information.

[0048]Next, in step ST3, the document feature extraction part 4 searches for the document vector over each document of the search-results document set 3. As shown in drawing 3, based on the frequency of occurrence of each word for every document, a document vector, each document D1, D2, ..., Dj, and ... each word KW1 which receives without Dm, KW2, ..., KWi, ..., the statistical dignity Wij of KWn are computed, and the dignity of each word for every document is expressed in the form of a vector. What is necessary is for the calculating method of this statistical dignity to have various calculating methods, such as TF-IDF and chi square statistic, and to choose it suitably according to the purpose and just to use it. The document feature extraction part 4 outputs the document vector set 5 produced by computing statistical dignity (document feature extraction step).

[0049]Next, in step ST4, the subject classification part 6 computes the similarity between document vectors, and subject is created by classifying the document vector set 5 into two or more sets according to similarity (subject classification step). It is divided roughly into two kinds such as the document group which gives and classifies a classification category top-down as a method of classifying a document, and the clustering which collects the document similar bottom-up.

[0050]In a document group, the category of a classification place is set up beforehand, from the sample document belonging to a category, the statistical dignity to a classification category is learned as a classification category vector, and the similarity of each document vector of the inputted document vector set 5 and a classification category vector is computed. The inner product value of a vector is used for similarity, for example. Thus, a document is classified into a classification category with the highest similarity based on the obtained similarity. The document vectors with similarity high on the other hand which compute the similarity between all the document vectors which exist in the document vector set 5 as which clustering was inputted are collectively made into one cluster, and a document is classified by repeating calculation of the similarity to a cluster, and the processing to arrange.

[0051]The subject classification part 6 should just have the function to classify a document vector, and is not restricted to the document group and clustering which were mentioned above, and other classifying methods (for example, principal component analysis) may be used for it. The subject classification part 6 makes a search term a word with high importance computed for every subject by making the classified set into subject. For example, the frequency of occurrence of each word is counted, consider it as importance, and let top 1 constant of importance be a search term, after deleting the word which it considered that is unnecessary and was separately set up about the word whole sentence in the letter [belonging to a certain subject]. The subject classification part 6 outputs the document information belonging to each subject, and the search term set 7 with importance of each subject.

[0052]Next, in step ST5, when the user's 13 search is the first time, it progresses to step ST6, and when the user's 13 search is not the first time, it progresses to step ST7. In step ST6, the presentation search term candidate generation part 8 chooses the presentation search term candidate 12 from the search term set 7 with importance of each subject on a predetermined standard, and follows him to step ST8. With the predetermined standard for choosing the presentation search term candidate 12, top 1 constant is chosen, for example in order of importance. On the other hand, in step ST7, the narrowing effect indicator 11 computed by the narrowing-down effect estimating part 10 is given to each search term candidate, and the narrowing effect indicator 11 chooses a high search term candidate, and considers it as the presentation search term candidate 12. The presentation search term candidate generation part 8 outputs the presentation search term candidate 12 who did in this way and was chosen to the classification result presentation part 14 with the document information belonging to corresponding subject (presentation search term candidate generation step).

[0053]Next, in step ST8, the classification result presentation part 14 shows the user 13 the presentation search term candidate 12 as the classification result 15 which was matched with each subject and visualized with the narrowing effect indicator 11 (classification result presentation step). As shown, for example in drawing 4, a search term-subject conversion table is used for the method of visualization. Subject corresponding with the presentation search term candidate 12 is expressed by the form of the procession, and a search term-subject conversion table shows the user 13 the narrowing effect indicator 11 as information visualized by each element of the procession. The classification result presentation part 14 will enable it to refer to a list of the document belonging to the specified subject, and various kinds of document information, if the user 13 specifies each subject T1, T2, ..., any of T6 they are using the document information belonging to corresponding subject.

[0054]Next, in step ST9, when the user 13 specifies subject with reference to the search term-subject conversion table which the classification result presentation part 14 was shown, it progresses to step ST10, and when the user 13 does not specify subject, it progresses to step ST12. In step ST10, the user 13 narrows down the subject 18 to specify and inputs into the effect estimating part 10. Next, in step ST11, the search term set 9 with document information and importance belonging to each subject comes to hand with reference to the presentation search term candidate generation part 8, and the narrowing effect indicator 11 of the search term belonging to the subject 18 specified by the user 13 is computed (the narrowing-down effect estimating step). Statistical dignity with each subject is again computed about the search term which belongs to the subject 18 which the user 13 specified, for example as a calculating method of the narrowing effect indicator 11. Or let the document number containing the search term belonging to the subject 18 specified by the user 13 be the narrowing effect indicator 11. Thus, the computed narrowing effect indicator 11 is outputted to the presentation search term candidate generation part 8, It is outputted to the classification result presentation part 14 from the presentation search term candidate generation part 8, It visualizes, as shown in a search term-subject conversion table as the classification result presentation part 14 indicated to be to drawing 4, and the narrowing effect indicator 11 is shown to the user 13 by correspondence with the presentation search term candidate 12 and each subject belonging to the subject 18 specified by the user 13. When processing of step ST11 is completed, it returns to step ST7.

[0055]since the range of choice of subject spreads when it enables it to specify two or more subjects about the subject 18 inputted into the narrowing-down effect estimating part 10, the contents of the document for which the user 13 asks can be specified more exactly. For example, what is necessary is to choose the higher rank fixed number as order with high importance by the search term belonging to the subject by which plural specifications were carried out, and for correspondence with the selected presentation search term candidate 12 and each subject just to show the user 13 the narrowing effect indicator 11.

[0056]The tendency of the narrowing effect indicator 11 of the wider range can grasp now the narrowing effect indicator 11 which the narrowing-down effect estimating part 10 computes to the whole search results by computing not per search term unit but per subject. For example, what is necessary is to create the weight vector of subject, to ask for similarity with each subject as an inner product value of a vector, and just to show the user 13 the similarity of the specified subject and each subject according to the form of a procession based on the word frequency of occurrence in the document belonging to the specified subject.

[0057]On the other hand, in step ST12, when it points to reclassification, without the user 13 inputting the subject 18 to the narrowing-down effect estimating part 10 with reference to the search term-subject conversion table which the classification result presentation part 14 presented, it progresses to step ST13, and in not directing reclassification, it ends this processing.

[0058]In step ST13, the user 13 inputs the directions information 16 for specifying the subject and the search term of the contents near a document to look for to the classification result presentation part 14, and gives directions of reclassification. The classification result presentation part 14 outputs the document ID number of the document belonging to the specified subject, and information, including the search term etc. which were specified, to the document feature set part 20 as the user's 13 selection information 19. Next, in step ST14 the document feature set part 20, Based on the user's 13 selection information 19, the dignity to the document vector corresponding to the inputted document ID number or the specified search term is changed, and the changed document vector 21 is outputted to the subject classification part 6 (document feature setting step). Change of dignity changes the dignity of the document belonging to the specified subject, or a search term by adding the constant value set up beforehand, for example to dignity. When processing of step ST14 is completed, it returns to step ST4.

[0059]As mentioned above, the narrowing-down effect estimating part 10 which computes the narrowing effect indicator 11 of the search term which belongs to the subject 18 specified by the user 13 according to this Embodiment 1, It has the classification result presentation part 14 which visualizes the narrowing effect indicator 11 and is shown to the user 13, Since the narrowing effect indicator 11 corresponding to the presentation search term candidate 12 and each subject belonging to the subject 18 specified by the user 13 was shown to the user 13, Since the user 13 can make selection of an additional search term easy, the additional high search term of the narrowing-down effect can be chosen exactly, and the effect that the efficiency of narrowing retrieval becomes good is acquired.

[0060]The subject classification part 6 which creates subject by classifying the document vector set 5 into two or more sets according to the similarity between the document vectors computed based on the document vector set 5 according to this Embodiment 1, It has the classification result presentation part 14 which visualizes the narrowing effect indicator 11 of a search term for every subject as shown in a search term-subject conversion table, and is shown to the user 13, Since it narrows down in the subject unit which is a bigger unit than a document unit and the effect was shown, the list nature of the narrowing-down effect to the whole document of search results increases, and the effect that the contents of the whole search results can be grasped efficiently is acquired.

[0061]Since according to this Embodiment 1 it has the document feature set part 20 by which the user 13 changes a document vector based on the directions information 16 directed as feedback information and was made to carry out reclassification to first-time search results, Since the document of the purpose distributed between subjects was brought together in one subject, the effect that the efficiency of the narrowing retrieval to the target document becomes good is acquired.

[0062]Embodiment 2. drawing 5 is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 2. In drawing 5, since drawing 1 and identical codes show a same or considerable portion, they omit the explanation. 31 is a subject classification item acquisition part (subject classification item acquisition means) which extracts the word by which a document is characterized as a subject classification item from the document in the search-results document set 3, and learns the weight vector to a subject classification item with reference to the text relevant to the extracted subject classification item.

32 is the subject classification item information that the subject classification item acquisition part 31 outputs to the subject classification part 6, and contains a subject classification item and a weight vector.

[0063]Next, operation is explained. In Embodiment 2, operation of the document set 1, the document retrieval part 2, the document feature extraction part 4, the presentation search term candidate generation part 8, the narrowing-down effect estimating part 10, the user 13, the classification result presentation part 14, and document feature set part 20 grade, And about the effect that these do so, since it is the same as that of Embodiment 1, the explanation is omitted.

[0064]The subject classification item acquisition part 31 extracts the word by which a document is characterized as a subject classification item from the document in the search-results document set 3, and the text relevant to the extracted subject classification item is referred to, The weight vector to a subject classification item is learned, and with a weight vector, a subject classification item is made into the subject classification item information 32, and is outputted to the subject classification part 6 (subject classification item acquisition step).

[0065]It is a domain name etc. which are contained in the title and URL of the HTML document obtained in a document with reference to a word with high incidence, and the HTML tag described by the HTML document as a subject classification item, for example. The text relevant to the extracted subject classification item, For example, the text relevant to a subject classification item is extracted by copying the text which analyzed the text of the position circumference in which a subject classification item exists, detected specific HTML tags, such as a paragraph pause, ******, an itemized statement, and a link destination, and was related with the detected HTML tag.

[0066]Using the text relevant to the extracted subject classification item, the subject classification item acquisition part 31 learns the weight vector to a subject classification item, and outputs a subject classification item to the subject classification part 6 with a weight vector. The subject classification part 6 classifies a document into a subject classification item with the highest similarity by computing the similarity of the subject classification item information 32 that it inputted from the subject classification item acquisition part 31, and each document vector of the document vector set 5 inputted from the document feature extraction part 4, for example with the inner product value of a vector.

[0067]As mentioned above, according to this Embodiment 2, do so the same effect as Embodiment 1, and. Extract a subject classification item from the document in the search-results document set 3, and it has the subject classification item acquisition part 31 outputted with the weight vector to the subject classification item concerned, Since the document was classified based on the similarity of the subject classification item information 32 and each document vector, Since it becomes unnecessary for the subject classification item used by the subject classification part 6 to be able to acquire automatically, and to set up a classification category beforehand, setting out of the document item of search results becomes unnecessary, and the effect that the efficiency of the user's 13 narrowing retrieval becomes good is acquired.

[0068]Embodiment 3. drawing 6 is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 3. In drawing 6, since drawing 1 and identical codes show a same or considerable portion, they omit the explanation. The specified document feature extraction part (specified document feature extraction means) which computes the document in which 41 was specified by the user 13 to a document vector, the document vector by which the specified document feature extraction part 41 outputs 42 to the document feature set part 20, and 43 are specified documents which the user 13 outputs to the specified document feature extraction part 41.

[0069]Next, operation is explained. In Embodiment 3, about the effect that operation of the document set 1, the document retrieval part 2, the document feature extraction part 4, the subject classification part 6, the presentation search term candidate generation part 8, the narrowing-down effect estimating part 10, and classification result presentation part 14 grade and these do so, since it is the same as that of Embodiment 1, the explanation is omitted.

[0070]The user 13 chooses the document of the contents near the document for which the user 13 asks with reference to the classification result presentation part 14, and directs the document 43 specified as the specified document feature extraction part 41. The specified document feature extraction part 41 computes the document vector 42 by calculating statistical dignity based on the appearance frequency of the word contained in the specified document 43, and outputs it to the document feature set part 20. The document feature set part 20 calculates the similarity of the document vector 42 of the specified document 43, and the document vector set 5 inputted from the document feature extraction part 4, and changes the dignity of the document vector in the document vector set 5 (specified document feature extraction step). For example, the document vector of the higher rank fixed number is chosen as the high order of similarity from the document vector set 5, and similarity is added to the dignity of a document vector and is changed. The subject classification part 6 classifies to the changed document vector set 5.

[0071]As mentioned above, according to this Embodiment 3, do so the same effect as Embodiment 1, and. It has the specified document feature extraction part 41 which computes a document vector from the document specified by the user 13, Since the similarity of the document vector 42 which the specified document feature extraction part 41 outputs, and the document vector set 5 which the document feature extraction part 4 outputted is calculated and the dignity of the document vector was changed, By specifying directly the document 43 of the contents near the document for which the user 13 asks, change of the document vector of the document feature set part 20 is attained, and the effect that the efficiency of the user's 13 narrowing retrieval becomes good is acquired.

[0072]Embodiment 4. drawing 7 is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 4. In drawing 7, since drawing 1 and identical codes show a same or considerable portion, they omit the explanation. The related term dictionary in which the related term relevant to a word and the word concerned in 51 is defined (the 1st recording device), The related term set part which 52 will choose the related term relevant to a search term from the related term dictionary 51 if a search term is inputted, and outputs a related term (related term setting-out means), They are a search term which the document feature set part 20 outputs 53, and is inputted into the related term set part 52, and a related term which the related term set part 52 outputs 54, and is inputted into the document feature set part 20.

[0073]Drawing 8 is an explanatory view showing an example which defined the word and related term in this embodiment of the invention 4. In drawing 8, the related term makes a variant notation and similar words the related term, and corresponds with the word described by each line.

[0074]Next, operation is explained. In Embodiment 4, operation of the document set 1, the document retrieval part 2, the document feature extraction part 4, the subject classification part 6, the presentation search term candidate generation part 8, the narrowing-down effect estimating part 10, the user 13, and classification result presentation part 14 grade, And about the effect that these do so, since it is the same as that of Embodiment 1, the explanation is omitted.

[0075]As shown in drawing 8, the predetermined word and the word relevant to the word concerned are beforehand recorded on the related term dictionary 51 (1st record step). The related term set part 52 extracts the variant notation and similar words corresponding to the inputted search term 53 with reference to the related term dictionary 51, considers it as the related term 54, and outputs the related term 54 to the document feature set part 20 (related term setting step). The document feature set part 20 adds the related term 54 inputted from the related term set part 52 to the selection information 19 of the user 13 who inputted from the classification result presentation part 14, and changes the document vector set 5 inputted from the document feature extraction part 4.

[0076]As mentioned above, according to this Embodiment 4, do so the same effect as Embodiment 1, and. It has the related term set part 52 which outputs the related term 54 with reference to the related term dictionary 51, Since the related term 54 inputted from the related term set part 52 is added to the selection information 19 of the user 13 who inputted from the classification result presentation part 14 and the document vector set 5 was changed, Since a document including a variant notation, similar words, etc. of the search term 53 can be searched, retrieval omission is controlled, and the effect that the efficiency of the search for discovering the document for which the user 13 asks becomes good is acquired.

[0077]Embodiment 5. drawing 9 is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 5. In drawing 9, since drawing 1 and identical codes show a same or considerable portion, they omit the explanation. The retrieval-required creation knowledge which 61 comprises a magnetic recording medium etc. and records the creation knowledge of retrieval required (the 2nd recording device), The retrieval-required preparing part which 62 will choose the optimal retrieval request statement based on a retrieval-required search term from the retrieval-required creation knowledge 61 if a retrieval-required search term is inputted, and outputs a retrieval request statement (retrieval-required preparing means), They are a retrieval-required search term which the presentation search term candidate generation part 8 outputs 63, and is inputted into the retrieval-required preparing part 62, and a retrieval request statement which the retrieval-required preparing part 62 outputs 64, and is inputted into the document retrieval part 2.

[0078]Next, operation is explained. In Embodiment 5, about the effect that operation of the document set 1, the document feature extraction part 4, the subject classification part 6, the narrowing-down effect estimating part 10, the user 13, the classification result presentation part 14, and document feature set part 20 grade and these do so, since it is the same as that of Embodiment 1, the explanation is omitted.

[0079]The presentation search term candidate generation part 8 outputs the high search term of the narrowing effect indicator 11 to the retrieval-required preparing part 62 as the retrieval-required search term 63 with reference to the narrowing effect indicator 11 inputted from the narrowing-down effect estimating part 10. The retrieval-required search term 63 may be plural. The knowledge for creating the retrieval request statement 64 which directs execution of retrieval processing to the document retrieval part 2 is beforehand defined as the retrieval-required creation knowledge 61 (2nd record step). For example, the retrieval request statement 64 comprises two kinds such as a retrieving instruction and a search condition. As a retrieving instruction, the kind of commands, such as <retrieval execution>, <run state acquisition>, and <database specification>, is defined, for example. As a search condition, the symbolic convention of parameters, such as a logical operator between search terms and specification of the information acquired as search results, is defined.

[0080]The retrieval-required preparing part 62 sets the retrieval-required search term 63 as a search condition with reference to the retrieval-required creation knowledge 61 according to the definition of the retrieval request statement 64 (retrieval-required creation step). The retrieving instruction of the retrieval request statement 64 is made into <retrieval execution>, for example, creates the retrieval request statement 64, and outputs it to the document retrieval part 2. When setting up the search condition of the retrieval request statement 64, An AND operator describes as opposed to two or more retrieval-required search terms 63, and if it is beyond the threshold which the narrowing effect indicator 11 given to the retrieval-required search term 63 set up beforehand, it can consider that the narrowing-down effect is high, and an OR operation child can also describe so that a wide range document may be searched.

[0081]As mentioned above, according to this Embodiment 5, do so the same effect as Embodiment 1, and. Since it was made to search again based on the retrieval request statement 64 which is provided with the retrieval-required preparing part 62 which creates the retrieval request statement 64 with reference to the retrieval-required creation knowledge 61, and the retrieval-required preparing part 62 outputs, The document for which the user 13 who was not contained in first-time search results asks can be automatically searched from the document set 1, and the effect that the efficiency of the user's 13 narrowing retrieval becomes good is acquired.

[0082]

[Effect of the Invention]As mentioned above, the document retrieval means which searches and outputs the document which suits a search condition from a document set according to this invention, The document feature extraction means which outputs the document vector set which computes the statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of the word described by the document, The subject sorting means which outputs the document information and the search term with importance of subject which create subject and belong to subject by classifying a document vector set according to the similarity between document vectors, The narrowing-down effect estimation means which computes and outputs the narrowing effect indicator of the search term which belongs to subject with reference to document information and a search term with importance, The presentation search term candidate creating means which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs the document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, The classification result presenting means urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Since it constituted so that it might have a document feature setting means which changes and outputs the document vector included in a document vector set based on selection information, Since the narrowing effect indicator was given to the presentation search term candidate who shows for narrowing retrieval, selection of an additional search term can be made easy and the effect that narrowing retrieval can be performed efficiently is done so.

[0083]Since according to this invention it constituted so that the narrowing-down effect estimation means might compute and output a narrowing effect indicator to one or more subjects at least, Since narrowing retrieval can be performed in the subject unit which is a bigger unit than a document unit, the tendency of a narrowing effect indicator to the search results of the wider range can be grasped, and the effect that the contents of the document for which a user asks can be specified exactly is done so.

[0084]Since according to this invention the classification result presenting means constituted so that the presentation search term candidate belonging to subject and the subject concerned might be shown in the form of a procession and a narrowing effect indicator might be shown to each element of a procession, The list nature of the narrowing-down effect to the whole document of search results increases, and the effect that the contents of the whole search results can be grasped efficiently is done so.

[0085]According to this invention, extract the word by which the document concerned is characterized as a subject classification item from the document which a document retrieval means outputs, and the text relevant to a subject classification item is referred to, Learn the weight vector to a subject classification item, have a subject classification item acquisition means which outputs a subject classification item and a weight vector to a subject sorting means, and a subject sorting means, Since it constituted so that subject might be created by classifying a document vector set based on a subject classification item and a weight vector, Since it becomes unnecessary for the subject classification item used by a subject sorting means to be able to acquire automatically, and to set up a classification category beforehand, setting out of the document item of search results becomes unnecessary, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0086]According to this invention, since the subject classification item acquisition means constituted so that a subject classification item might be extracted from the document outputted from a document retrieval means based on the frequency of occurrence of the word described by the document concerned, the effect that a subject classification item can be extracted efficiently automatically is done so.

[0087]Since according to this invention the subject classification item acquisition means constituted so that a subject classification item might be extracted from the document outputted from a document retrieval means with reference to the tag described by the document concerned, The effect that a subject classification item can be automatically extracted from the document in which the tag is described efficiently is done so.

[0088]According to this invention, a document vector is computed from the document specified via the classification result presenting means, The document vector which is equipped with the specified document feature extraction means outputted to a document feature setting means and to which a document feature setting means is outputted from a specified document feature extraction means, Since it constituted so that a document vector set might be changed based on the document vector set outputted from a document feature extraction means, By specifying directly the document of the contents near the document for which a user asks, change of the document vector of a document feature setting means is attained, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0089]The 1st recording device that according to this invention defines the word relevant to a predetermined word and is recorded as a related term, It has a related term setting-out means to extract the related term corresponding to the specified search term from the 1st recording device, and to output to a document feature setting means, Since the document feature setting means constituted based on the selection information inputted from the related term and the classification result presenting means so that a document vector set might be changed, Since a document including a variant notation, similar words, etc. of a search term can be searched, retrieval omission is controlled, and the effect that the efficiency of the search for discovering the document for which a user asks becomes good is done so.

[0090]The 2nd recording device that records the creation knowledge of a retrieval request statement according to this invention, With reference to the 2nd recording device concerned, the retrieval request statement corresponding to the search term which the presentation search term candidate creating means outputted is created, Since it had the retrieval-required preparing means outputted to a document retrieval means, and the presentation search term candidate creating means constituted so that the search term outputted to a retrieval-required preparing means based on a narrowing effect indicator might be chosen, The document for which the user who was not contained in first-time search results asks can be automatically searched from a document set, and the effect that the efficiency of a user's narrowing retrieval becomes good is done so.

[0091]Since according to this invention the presentation search term candidate creating means constituted so that two or more search terms might be chosen, it might output to a retrieval-required preparing means and a retrieval-required preparing means might create a retrieval request statement from the logical operation to two or more search terms, By an AND operation, narrowing retrieval can be performed more exactly and the effect that narrowing retrieval can be performed more to a large area is done so by an OR operation.

[0092]The document-retrieval step which searches and outputs the document which suits a search condition from a document set according to this invention, The document feature extraction step which outputs the document vector set which computes the statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of the word described by the document, The subject classification step which outputs the document information and the search term with importance of subject which create subject and belong to subject by classifying a document vector set according to the similarity between document vectors, The narrowing-down effect estimating step which computes and outputs the narrowing effect indicator of the search term which belongs to subject with reference to document information and a search term with importance, The presentation search term candidate generation step which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs the document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, The classification result presentation step urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Since it constituted so that it might have the document feature setting step which changes and outputs the document vector included in a document vector set based on selection information, Since the narrowing effect indicator was given to the presentation search term candidate who shows for narrowing retrieval, selection of an additional search term can be made easy and the effect that narrowing retrieval can be performed efficiently is done so.

[0093]Since according to this invention it constituted so that the narrowing-down effect estimating step might compute and output a narrowing effect indicator to one or more subjects at least, Since narrowing retrieval can be performed in the subject unit which is a bigger unit than a document unit, the tendency of a narrowing effect indicator to the search results of the wider range can be grasped, and the effect that the contents of the document for which a user asks can be specified exactly is done so.

[0094]Since according to this invention the classification result presentation step constituted so that the presentation search term candidate belonging to subject and the subject concerned might be shown in the form of a procession and a narrowing effect indicator might be shown to each element of a procession, The list nature of the narrowing-down effect to the whole document of search results increases, and the effect that the contents of the whole search results can be grasped efficiently is done so.

[0095]According to this invention, extract the word by which the document concerned is characterized as a subject classification item from the document which a document-retrieval step outputs, and the text relevant to a subject classification item is referred to, Learn the weight vector to a subject classification item, and it has a subject classification item acquisition step which outputs a subject classification item and a weight vector, Since the subject classification step constituted so that subject might be created by classifying a document vector set based on a subject classification item and a weight vector, Since it becomes unnecessary for the subject classification item used by a subject classification step to be able to acquire automatically, and to set up a classification category beforehand, setting out of the document item of search results becomes unnecessary, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0096]According to this invention, since the subject classification item acquisition step constituted so that a subject classification item might be extracted from the document outputted from a document-retrieval step based on the frequency of occurrence of the word described by the document concerned, the effect that a subject classification item can be extracted efficiently automatically is done so.

[0097]Since according to this invention the subject classification item acquisition step constituted so that a subject classification item might be extracted from the document outputted from a document-retrieval step with reference to the tag described by the document concerned, The effect that a subject classification item can be automatically extracted from the document in which the tag is described efficiently is done so.

[0098]The document vector which has a specified document feature extraction step which computes and outputs a document vector from the document specified via the classification result presentation step according to this invention and to which the document feature setting step was outputted from the specified document feature extraction step, Since it constituted so that a document vector set might be changed based on the document vector set outputted from the document feature extraction step, By specifying directly the document of the contents near the document for which a user asks, change of the document vector of the document feature setting step is attained, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0099]The 1st record step that according to this invention defines the word relevant to a predetermined word and is recorded as a related term, It has a related term setting step which extracts and outputs the related term corresponding to the specified search term, Since the document feature setting step constituted based on the selection information inputted from the related term and the classification result presentation step so that a document vector set might be changed, Since a document including a variant notation, similar words, etc. of a search term can be searched, retrieval omission is controlled, and the effect that the efficiency of the search for discovering the document for which a user asks becomes good is done so.

[0100]The 2nd record step that records the creation knowledge of a retrieval request statement according to this invention, It has a retrieval-required creation step which creates and outputs the retrieval request statement corresponding to the search term which the presentation search term candidate generation step outputted, Since the presentation search term candidate generation step constituted so that the search term outputted based on a narrowing effect indicator might be chosen, The document for which the user who was not contained in first-time search results asks can be automatically searched from a document set, and the effect that the efficiency of a user's narrowing retrieval becomes good is done so.

[0101]Since according to this invention the presentation search term candidate generation step chose and outputted two or more search terms, and the retrieval-required creation step constituted so that a retrieval request statement might be created from the logical operation to two or more search terms, By an AND operation, narrowing retrieval can be performed more exactly and the effect that narrowing retrieval can be performed more to a large area is done so by an OR operation.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Field of the Invention]This invention relates to the document retrieval system and document retrieval method which can perform a search of a document efficiently by [which is a bigger unit than a document] narrowing down for every subject, showing an effect and making selection of an additional search term easy.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Description of the Prior Art]When the electronic document etc. which are recorded and managed are searched to the HTML document which can be perused using the Internet, or a large-scale text database, The amount of information of the document information acquired as search results increases dramatically, and many time and labors are needed for discovery of the document for which a user asks. For this reason, according to the contents, classify the document information acquired as search results so that a user can search the target document efficiently, or. The demand to the art which supports discovery of the document for which a user who shows a user the search term candidate who adds for narrowing retrieval asks is increasing.

[0003]About the art in which a user can search the target document efficiently, JP,H9-231238,A "text browsing result display method and device." There are (document 1 being called hereafter) and art indicated by JP,H11-213000,A "storage which stored the interactive information retrieval method, the device, and the interactive information retrieval program" (document 2 is called hereafter).

[0004]The art indicated in document 1 carries out the theme classification of the search results by fuzzy clustering, and shows a user the category by which the theme classification was carried out with the group of a search term. The art indicated in document 2 classifies search results according to clustering, shows a user the classified category with the group of a search term, classifies the category specified by a user into a subcategory further, and enables interactive search.

[0005]being related with the art which supports narrowing retrieval -- "development of the interface for search-results narrowing down in a WWW search service" (Information Processing Society of Japan and a human interface study group (HI76-5).) of Hayakawa and others In pp.25-1998 and the following, there is art which set to call document 3 and was indicated.

[0006]The art indicated in document 3 provides the interface which visualizes the information how many search-results numbers a search term can narrow down, and is shown to a user, in order to be able to grasp easily the relation between a search term and a search-results document set. The narrowing-down result of a search term is visualized using a document word matrix, The matrix which has a word and document in a row and column, respectively is shown to a direct user in the form of a table, the dignity of the word to document is expressed with the luminosity of the cell of a matrix, and it enables it to look through the narrowing-down effect of an additional search term.

[0007]There is art indicated by JP,H11-85764,A "storage which stored the statistical estimation program of the statistical estimation method of the search-results number, the device, and the search-results number" (document 4 is called hereafter) about the conventional technology described in document 3. The art indicated in document 4 is indicated about the statistical estimation method of the search-results number for knowing how the number of search results will change, when the search term for narrowing down search results further in document 3 is added.

[0008]Drawing 10 is a lineblock diagram showing the conventional document retrieval system. It is a lineblock diagram of the estimating device which is an example of the operation indicated by document 4.

An estimating device for the document retrieval according [on drawing 10 and / 101] to a full-text search, A search term for 102 to search document, the text database with which document is recorded and managed 103, The search-results total of document in which 104 is outputted from the text database 103, The document sample set which 105 extracts 50 affairs at random out of search results, and is outputted from the text database 103, The document-word-matrix generation part in which the document sample set 105 is inputted into and 106 counts the appearance frequency of each word about each document, and 107 are the document word matrices which the document-word-matrix generation part 106 generated. The document sample set 105 is not limited to 50 affairs, and can be suitably set up according to a situation.

[0009]The search term candidate presentation part in which 108 calculates word significance to the document word matrix 107 in drawing 10, The search term candidate to whom 109 is outputted from the search term candidate presentation part 108, the search term selecting part by which 110 outputs the search term candidate 109 to a monitor etc., The selection signal into which 111 is inputted when a user chooses the search term candidate 109, The additional search term to which 112 is outputted from the search term selecting part 110 based on the selection signal 111, The number-of-cases estimating part in which 113 calculates the incidence of the additional search term 112 based on the document word matrix 107 and the additional search term 112, and 114 are the presumed numbers produced by being outputted from the number-of-cases estimating part 113, and multiplying the search-results total 104 by the incidence. Drawing 11 is an explanatory view showing an example of the document word matrix 107 in the conventional document retrieval system.

[0010]Next, operation is explained. If the search term 102 inputs into the text database 103, based on the search term 102, document recorded and managed will be searched to the text database 103. The text database 103 extracts 50 affairs at random out of search results, and outputs them to the document-word-matrix generation part 106 as the document sample set 105, and it outputs the search-results total 104 to the number-of-cases estimating part 113. By counting the number of times to which each word appears about each document based on the document sample set 105, the document-word-matrix generation part 106 generates the document word matrix 107 shown in drawing 11, and outputs it to the search term candidate presentation part 108 and the number-of-cases estimating part 113. In the document word matrix 107, a row and column shows a literature identifier and a search term list, respectively, and the value of the cell of a table shows the number of times to which the search term of a sequence [be / it / under / of corresponding document of a line / correspondence] appears.

[0011]The search term candidate presentation part 108 calculates the word significance which shows how important words arbitrary about arbitrary document are except for words, such as a particle and a pronoun, from the document word matrix 107. Word significance is an index which becomes high and becomes low with the word which has appeared by many document conversely with the word which has appeared in specific document intensively. Based on the calculated word significance, word significance outputs the search term candidate presentation part 108 to the search term selecting part 110 by making a high word into the search term candidate 109.

[0012]The search term selecting part 110 outputs the search term candidate 109 to a monitor etc., makes a user choose arbitrary search terms from the search term candidate 109, and is made to input as the selection signal 111. The search term selecting part 110 is outputted to the number-of-cases estimating part 113 as the additional search term 112 based on the selection signal 111 which the user inputted.

[0013]The search-results total 104, the document word matrix 107, and the additional search term 112 input into the number-of-cases estimating part 113. The number-of-cases estimating part 113 counts the number of the lines whose sequences of the additional search term 112 of the document word matrix 107 are not "0", and the incidence of the additional search term 112 is obtained by **(ing) this with the total number of lines. The number-of-cases estimating part 113 outputs the presumed number 114 produced by multiplying the incidence of the additional search term 112 by the search-results total 104.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Effect of the Invention]As mentioned above, the document retrieval means which searches and outputs the document which suits a search condition from a document set according to this invention, The document feature extraction means which outputs the document vector set which computes the statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of the word described by the document, The subject sorting means which outputs the document information and the search term with importance of subject which create subject and belong to subject by classifying a document vector set according to the similarity between document vectors, The narrowing-down effect estimation means which computes and outputs the narrowing effect indicator of the search term which belongs to subject with reference to document information and a search term with importance, The presentation search term candidate creating means which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs the document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, The classification result presenting means urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Since it constituted so that it might have a document feature setting means which changes and outputs the document vector included in a document vector set based on selection information, Since the narrowing effect indicator was given to the presentation search term candidate who shows for narrowing retrieval, selection of an additional search term can be made easy and the effect that narrowing retrieval can be performed efficiently is done so.
[0083]Since according to this invention it constituted so that the narrowing-down effect estimation means might compute and output a narrowing effect indicator to one or more subjects at least, Since narrowing retrieval can be performed in the subject unit which is a bigger unit than a document unit, the tendency of a narrowing effect indicater to the search results of the wider range can be grasped, and the effect that the contents of the document for

which a user asks can be specified exactly is done so.

[0084]Since according to this invention the classification result presenting means constituted so that the presentation search term candidate belonging to subject and the subject concerned might be shown in the form of a procession and a narrowing effect indicator might be shown to each element of a procession, The list nature of the narrowing-down effect to the whole document of search results increases, and the effect that the contents of the whole search results can be grasped efficiently is done so.

[0085]According to this invention, extract the word by which the document concerned is characterized as a subject classification item from the document which a document retrieval means outputs, and the text relevant to a subject classification item is referred to, Learn the weight vector to a subject classification item, have a subject classification item acquisition means which outputs a subject classification item and a weight vector to a subject sorting means, and a subject sorting means, Since it constituted so that subject might be created by classifying a document vector set based on a subject classification item and a weight vector, Since it becomes unnecessary for the subject classification item used by a subject sorting means to be able to acquire automatically, and to set up a classification category beforehand, setting out of the document item of search results becomes unnecessary, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0086]According to this invention, since the subject classification item acquisition means constituted so that a subject classification item might be extracted from the document outputted from a document retrieval means based on the frequency of occurrence of the word described by the document concerned, the effect that a subject classification item can be extracted efficiently automatically is done so.

[0087]Since according to this invention the subject classification item acquisition means constituted so that a subject classification item might be extracted from the document outputted from a document retrieval means with reference to the tag described by the document concerned, The effect that a subject classification item can be automatically extracted from the document in which the tag is described efficiently is done so.

[0088]According to this invention, a document vector is computed from the document specified via the classification result presenting means, The document vector which is equipped with the specified document feature extraction means outputted to a document feature setting means and to which a document feature setting means is outputted from a specified document feature extraction means, Since it constituted so that a document vector set might be changed based on the document vector set outputted from a document feature extraction means, By specifying directly the document of the contents near the document for which a user asks, change of the document vector of a document feature setting means is attained, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0089]The 1st recording device that according to this invention defines the word relevant to a predetermined word and is recorded as a related term, It has a related term setting-out means to extract the related term corresponding to the specified search term from the 1st recording device, and to output to a document feature setting means, Since the document feature setting means constituted based on the selection information inputted from the related term and the classification result presenting means so that a document vector set might be changed, Since a document including a variant notation, similar words, etc. of a search term can be searched, retrieval omission is controlled, and the effect that the efficiency of the search for discovering the document for which a user asks becomes good is done so.

[0090]The 2nd recording device that records the creation knowledge of a retrieval request statement according to this invention, With reference to the 2nd recording device concerned, the retrieval request statement corresponding to the search term which the presentation search term candidate creating means outputted is created, Since it had the retrieval-required preparing means outputted to a document retrieval means, and the presentation search term candidate creating means constituted so that the search term outputted to a retrieval-required preparing means based on a narrowing effect indicator might be chosen, The document for which the user who was not contained in first-time search results asks can be automatically searched from a document set, and the effect that the efficiency of a user's narrowing retrieval becomes good is done so.

[0091]Since according to this invention the presentation search term candidate creating means constituted so that two or more search terms might be chosen, it might output to a retrieval-required preparing means and a retrieval-required preparing means might create a retrieval request statement from the logical operation to two or more search terms, By an AND operation, narrowing retrieval can be performed more exactly and the effect that narrowing retrieval can be performed more to a large area is done so by an OR operation.

[0092]The document-retrieval step which searches and outputs the document which suits a search condition from a document set according to this invention, The document feature extraction step which outputs the document vector set which computes the statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of the word described by the document, The subject classification step which outputs the document information and the search term with importance of subject which create subject and belong to subject by classifying a document vector set according to the similarity between document vectors, The narrowing-down effect estimating step which computes and outputs the narrowing effect indicator of the search term which belongs to subject with reference to document information and a search term with importance, The presentation search term candidate generation step which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs the document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, The classification result presentation step urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Since it constituted so that it might have the document feature setting step which changes and outputs the document vector included in a document vector set based on selection information, Since the narrowing effect indicator was given to the presentation search term candidate who shows for narrowing retrieval, selection of an additional search term can be made easy and the effect that narrowing retrieval can be performed efficiently is done so.

[0093]Since according to this invention it constituted so that the narrowing-down effect estimating step might compute and output a narrowing effect indicator to one or more subjects at least, Since narrowing retrieval can be performed in the subject unit which is a bigger unit than a document unit, the tendency of a narrowing effect indicator to the search results of the wider range can be grasped, and the effect that the contents of the document for which a user asks can be specified exactly is done so.

[0094]Since according to this invention the classification result presentation step constituted so that the presentation search term candidate belonging to subject and the subject concerned might be shown in the form of a procession and a narrowing effect indicator might be shown to each element of a procession, The list nature of the narrowing-down effect to the whole document of search results increases, and the effect that the contents of the whole search results can be grasped efficiently is done so.

[0095]According to this invention, extract the word by which the document concerned is characterized as a subject classification item from the document which a document-retrieval step outputs, and the text relevant to a subject classification item is referred to, Learn the weight vector to a subject classification item, and it has a subject classification item acquisition step which outputs a subject classification item and a weight vector, Since the subject classification step constituted so that subject might be created by classifying a document vector set based on a subject classification item and a weight vector, Since it becomes unnecessary for the subject classification item used by a subject classification step to be able to acquire automatically, and to set up a classification category beforehand, setting out of the document item of search results becomes unnecessary, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0096]According to this invention, since the subject classification item acquisition step constituted so that a subject classification item might be extracted from the document outputted from a document-retrieval step based on the frequency of occurrence of the word described by the document concerned, the effect that a subject classification item can be extracted efficiently automatically is done so.

[0097]Since according to this invention the subject classification item acquisition step constituted so that a subject classification item might be extracted from the document outputted from a document-retrieval step with reference to the tag described by the document concerned, The effect that a subject classification item can be automatically extracted from the document in which the tag is described efficiently is done so.

[0098]The document vector which has a specified document feature extraction step which computes and outputs a document vector from the document specified via the classification result presentation step according to this invention and to which the document feature setting step was outputted from the specified document feature extraction step, Since it constituted so that a document vector set might be changed based on the document vector set outputted from the document feature extraction step, By specifying directly the document of the contents near the document for which a user asks, change of the document vector of the document feature setting step is attained, and the effect that the efficiency of narrowing retrieval becomes good is done so.

[0099]The 1st record step that according to this invention defines the word relevant to a predetermined word and is recorded as a related term, It has a related term setting step which extracts and outputs the related term corresponding to the specified search term, Since the document feature setting step constituted based on the selection information inputted from the related term and the classification result presentation step so that a document vector set might be changed, Since a document including a variant notation, similar words, etc. of a search term can be searched, retrieval omission is controlled, and the effect that the efficiency of the search for discovering the document for which a user asks becomes good is done so.

[0100]The 2nd record step that records the creation knowledge of a retrieval request statement according to this invention, It has a retrieval-required creation step which creates and outputs the retrieval request statement corresponding to the search term which the presentation search term candidate generation step outputted, Since the presentation search term candidate generation step constituted so that the search term outputted based on a narrowing effect indicator might be chosen, The document for which the user who was not contained in first-time search results asks can be automatically searched from a document set, and the effect that the efficiency of a user's narrowing retrieval becomes good is done so.

[0101]Since according to this invention the presentation search term candidate generation step chose and outputted two or more search terms, and the retrieval-required creation step constituted so that a retrieval request statement might be created from the logical operation to two or more search terms, By an AND operation, narrowing retrieval can be performed more exactly and the effect that narrowing retrieval can be performed more to a large area is done so by an OR operation.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Problem(s) to be Solved by the Invention]Since the conventional document retrieval system is constituted as mentioned above, about the narrowing-down effect of a search term, according to the conventional technology indicated by document 1 and document 2, classified the document of search results, and have shown the user the search term in the classified category by it, but. Since there was no information on the narrowing-down effect corresponding to the shown search term when choosing in order to use the shown search term for narrowing retrieval, SUBJECT that it was difficult to choose the search term for carrying out narrowing retrieval efficiently occurred.

[0015]Although the conventional document retrieval system has shown the user the narrowing-down effect of a search term per document by the conventional technology indicated by document 3 about visualization of search results, When search results became a scale which is about thousands, the display of the narrowing-down effect of the whole search results by a document word matrix became very difficult, and SUBJECT that list nature was missing occurred.

[0016]By the conventional technology indicated by document 4, the conventional document retrieval system about visualization of search results. From the incidence of the additional search term which outputs the document sample set from search results, and corresponds for every document based on the document sample set, since the narrowing-down number to the whole search results is presumed, Since it became the presentation for every additional search term when showing a user the presumed narrowing-down number, the display of the narrowing-down effect of the whole search results became very difficult, and SUBJECT that list nature was missing occurred.

[0017]Although the conventional document retrieval system is carrying out specific classification of the category specified in order to narrow down to the target document to the subcategory by the conventional technology indicated by document 2 about the narrowing retrieval to the target document, The target document does not exist altogether in the specified category, Since the document for which a user asks existed in other categories, the document obtained by narrowing retrieval was limited only to the document which exists in the specified category and the retrieval omission of the target document arose, SUBJECT that it was difficult to perform narrowing retrieval again occurred.

[0018]By the conventional technology indicated by document 3 and document 4, the conventional document retrieval system about the narrowing retrieval to the target document. Since AND retrieval which adds the search term to first-time search results is performed, the retrieval object of narrowing retrieval, It was limited to the document of search results to the last search term, search of a document for it to be scattered to whole sentence document space became impossible, and SUBJECT that it was difficult to perform narrowing retrieval again occurred.

[0019]As opposed to the search term which it was made in order that this invention might solve above SUBJECT, and is shown for narrowing retrieval, The index showing the narrowing-down effect is given, selection of an additional search term is made easy, and it aims at acquiring the document retrieval system and document retrieval method which make a user's narrowing retrieval efficient.

[0020]This invention classifies the document of search results, and makes subject the classified document set, and the narrowing-down effect of a search term is shown to a user for every subject which is a bigger unit than a document, It aims at acquiring the document retrieval system and document retrieval method which improve the list nature of the narrowing-down effect to the whole document of search results.

[0021]When this invention carries out reclassification of the first-time search results using the subject which the user specified as feedback information, or the information on a search term, As the document of the purpose distributed between subjects is brought together in one subject, it aims at acquiring the document retrieval system and document retrieval method which make narrowing down of search results efficient.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

[Means for Solving the Problem]A document retrieval means which a document retrieval system concerning this invention searches a document which suits a search condition from a document set, and is outputted, A document feature extraction means which outputs a document vector set which computes statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of a word described by document, A subject sorting means which outputs document information and a search term with importance of subject which create subject and belong to subject by classifying a document vector set according to similarity between document vectors, The narrowing-down effect estimation means which computes and outputs a narrowing effect indicator of a search term which belongs to subject with reference to document information and a search term with importance, A presentation search term candidate creating means which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, A classification result presenting means urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Based on selection information, it has a document feature setting means which changes and outputs a document vector included in a document vector set.

[0023]The narrowing-down effect estimation means computes a narrowing effect indicator, and it is made to output a document retrieval system concerning this invention to one or more subjects at least.

[0024]A document retrieval system concerning this invention presents a presentation search term candidate to whom a classification result presenting means belongs to subject and the subject concerned in the form of a procession, and shows each element of a procession a narrowing effect indicator.

[0025]A document retrieval system concerning this invention extracts a word by which the document concerned is characterized as a subject classification item from a document which a document retrieval means outputs, and a text relevant to a subject classification item is referred to, Learn a weight vector to a subject classification item, have a subject classification item acquisition means which outputs a subject classification item and a weight vector to a subject sorting means, and a subject sorting means, Subject is created by classifying a document vector set based on a subject classification item and a weight vector.

[0026]A document retrieval system concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition means is outputted from a document retrieval means based on the frequency of occurrence of a word described by the document concerned.

[0027]A document retrieval system concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition means is outputted from a document retrieval means with reference to a tag described by the document concerned.

[0028]A document retrieval system concerning this invention computes a document vector from a document specified via a classification result presenting means, It has a specified document feature extraction means outputted to a document feature setting means, and a document feature setting means changes a document vector set based on a document vector outputted from a specified document feature extraction means, and a document vector set outputted from a document feature extraction means.

[0029]The 1st recording device that a document retrieval system concerning this invention defines a word relevant to a predetermined word, and is recorded as a related term, It has a related term setting-out means to extract a related term corresponding to a specified search term from the 1st recording device, and to output to a document feature setting means, and a document feature setting means changes a document vector set based on selection information inputted from a related term and a classification result presenting means.

[0030]The 2nd recording device on which a document retrieval system concerning this invention records creation knowledge of a retrieval request statement, With reference to the 2nd recording device concerned, a retrieval request statement corresponding to a search term which a presentation search term candidate creating means outputted is created, It has a retrieval-required preparing means outputted to a document retrieval means, and a presentation search term candidate creating means chooses a search term outputted to a retrieval-required preparing means based on a narrowing effect indicator.

[0031]A presentation search term candidate creating means chooses two or more search terms, and outputs a document retrieval system concerning this invention to a retrieval-required preparing means, and a retrieval-required preparing means creates a retrieval request statement from a logical operation to two or more search terms.

[0032]A document-retrieval step which a document retrieval method concerning this invention searches a document which suits a search condition from a document set, and is outputted, A document feature extraction step which outputs a document vector set which computes statistical dignity of a word and is obtained from dignity based on the frequency of occurrence of a word described by document, A subject classification step which outputs document information and a search term with importance of subject which create subject and belong to subject by classifying a document vector set according to similarity between document vectors, The narrowing-down effect estimating step which computes and outputs a narrowing effect indicator of a search term which belongs to subject with reference to document information and a search term with importance, A presentation search term candidate generation step which gives a narrowing effect indicator to a search term, a narrowing effect indicator chooses a high search term, considers it as a presentation search term candidate, and outputs document information corresponding to the presentation search term candidate concerned and the presentation search term candidate concerned, A classification result presentation step urged to match a presentation search term candidate and subject, to show with a narrowing effect indicator, and to input either directions information or selection information and both, Based on selection information, it has the document feature setting step which changes and outputs a document vector included in a document vector set.

[0033]The narrowing-down effect estimating step computes a narrowing effect indicator, and it is made to output a document retrieval method concerning this invention to one or more subjects at least.

[0034]A document retrieval method concerning this invention presents a presentation search term candidate to whom a classification result presentation step belongs to subject and the subject concerned in the form of a procession, and shows each element of a procession a narrowing effect indicator.

[0035]A document retrieval method concerning this invention extracts a word by which the document concerned is characterized as a subject classification item from a document which a document-retrieval step outputs, and a text relevant to a subject classification item is referred to, A weight vector to a subject classification item is learned, it has a subject classification item acquisition step which outputs a subject classification item and a weight vector, and a subject classification step creates subject by classifying a document vector set based on a subject classification item and a weight vector.

[0036]A document retrieval method concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition step is outputted from a document-retrieval step based on the frequency of occurrence of a word described by the document concerned.

[0037]A document retrieval method concerning this invention extracts a subject classification item from a document in which a subject classification item acquisition step is outputted from a document-retrieval step with reference to a tag described by the document concerned.

[0038]A document retrieval method concerning this invention has a specified document feature extraction step which computes and outputs a document vector from a document specified via a classification result presentation step, The document feature setting step changes a document vector set based on a document vector outputted from a specified document feature extraction step, and a document vector set outputted from a document feature extraction step.

[0039]The 1st record step that a document retrieval method concerning this invention defines a word relevant to a predetermined word, and is recorded as a related term, It has a related term setting step which extracts and outputs a related term corresponding to a specified search term, and a document vector set is changed based on selection information which the document feature setting step inputted from a related term and a classification result presentation step.

[0040]The 2nd record step on which a document retrieval method concerning this invention records creation knowledge of a retrieval request statement, It has a retrieval-required creation step which creates and outputs a retrieval request statement corresponding to a search term which a presentation search term candidate generation step outputted, and a presentation search term candidate generation step chooses a search term outputted based on a narrowing effect indicator.

[0041]A presentation search term candidate generation step chooses and outputs two or more search terms, and a document retrieval method concerning this invention creates a retrieval request statement from a logical operation [as opposed to two or more search terms in a retrieval-required creation step].

[0042]

[Embodiment of the Invention]Hereafter, one gestalt of implementation of this invention is explained.

Embodiment 1. drawing 1 is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 1. The HTML document which 1 is a document set used as a retrieval object in drawing 1, for example, can be perused using the Internet, Electronized texts, such as an electronic document recorded and managed, are consisted of by the E-mail which can be transmitted and received using the Internet, and the large-scale text database recorded on the recorder or the recording medium. The document retrieval part which searches a document from the document set 1 according to conditions predetermined in 2

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

[Brief Description of the Drawings]

[Drawing 1]It is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 1.

[Drawing 2]It is a flow chart explaining operation of the document retrieval system by this embodiment of the invention 1.

[Drawing 3]It is an explanatory view showing an example of the document vector in this embodiment of the invention 1.

[Drawing 4]It is an explanatory view showing an example of the search term-subject conversion table in this embodiment of the invention 1.

[Drawing 5]It is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 2.

[Drawing 6]It is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 3.

[Drawing 7]It is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 4.

[Drawing 8]It is an explanatory view showing an example which defined the word and related term in this embodiment of the invention 4.

[Drawing 9]It is a block diagram showing the composition of the document retrieval system by this embodiment of the invention 5.

[Drawing 10]It is a lineblock diagram showing the conventional document retrieval system.

[Drawing 11]It is an explanatory view showing an example of the document word matrix in the conventional document retrieval system.

[Description of Notations]

1 A document set and 2 A document retrieval part (document retrieval means) and 3 Search-results document set, 4 document feature extraction part (document feature extraction means) and 5 A document vector set and 6 Subject classification part (subject sorting means), 7 The search term set with importance, and 8 Presentation search term candidate generation part

(presentation search term candidate creating means), 9 The search term set with importance, the 10 narrowing-down effect estimating part (the narrowing-down effect estimation means), 11 A narrowing effect indicator and 12 A presentation search term candidate and 13 A user and 14 Classification result presentation part (classification result presenting means), 15 A classification result and 16 [Selection information,] Directions information and 17 A search condition and 18 Subject and 19 20 The document feature set part (document feature setting means) and 21 Document vector, 31 A subject classification item acquisition part (subject classification item acquisition means) and 32 Subject classification item information, 41 A specified document feature extraction part (specified document feature extraction means) and 42 Document vector, 43 A document, 51 related term dictionaries (the 1st recording device), 52 related-term set part (related term setting-out means), 53 search terms, and 54 [Retrieval request statement.] A related term and 61 Retrieval-required creation knowledge (the 2nd recording device), 62 retrieval-required preparing part (retrieval-required preparing means), and 63 A retrieval-required search term and 64

[Translation done.]

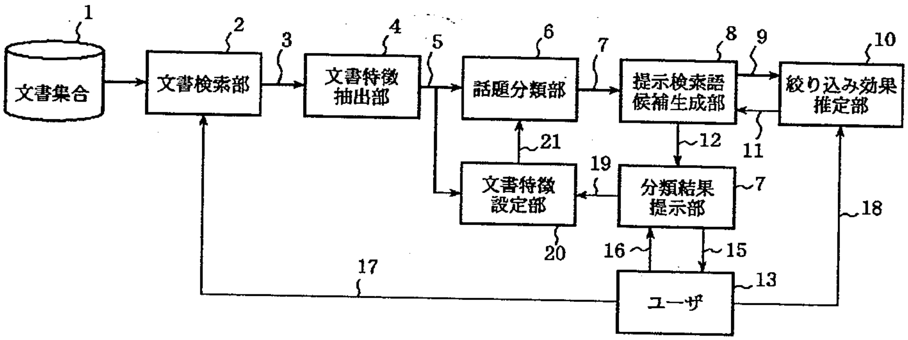
* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.

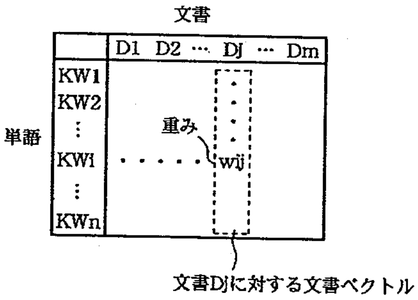
2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

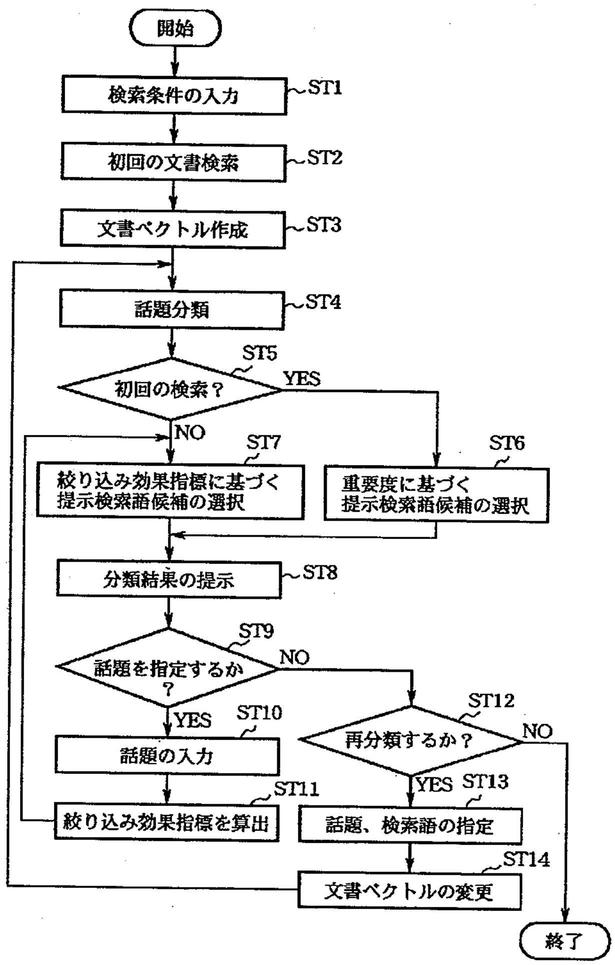
[Drawing 1]



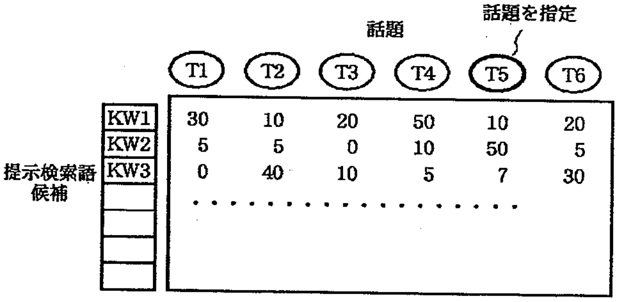
[Drawing 3]



[Drawing 2]

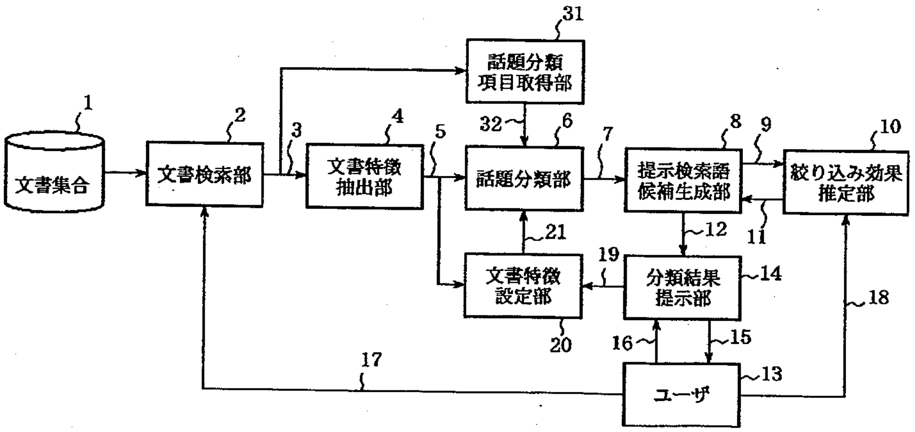


[Drawing 4]

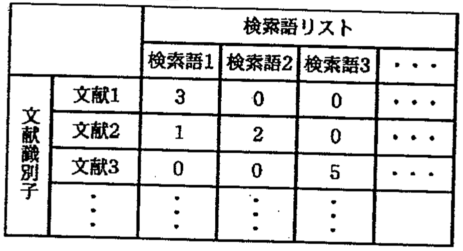

話題
話題を指定
提示検索語候補

	T1	T2	T3	T4	T5	T6
KW1	30	10	20	50	10	20
KW2	5	5	0	10	50	5
KW3	0	40	10	5	7	30
	・・・・・・・・・・・・・・・・・・					

[Drawing 5]

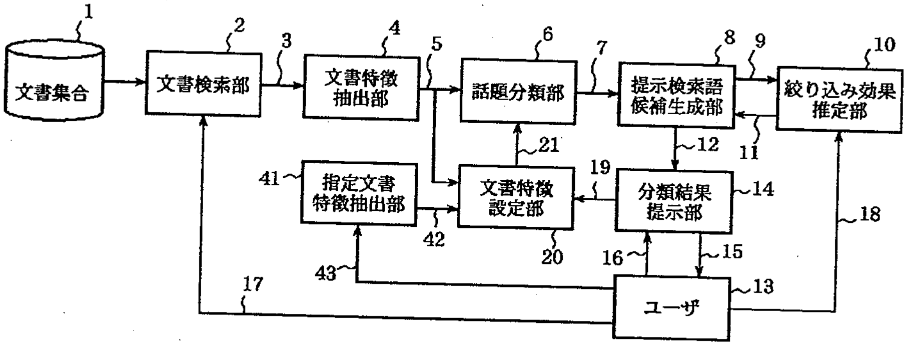


[Drawing 11]

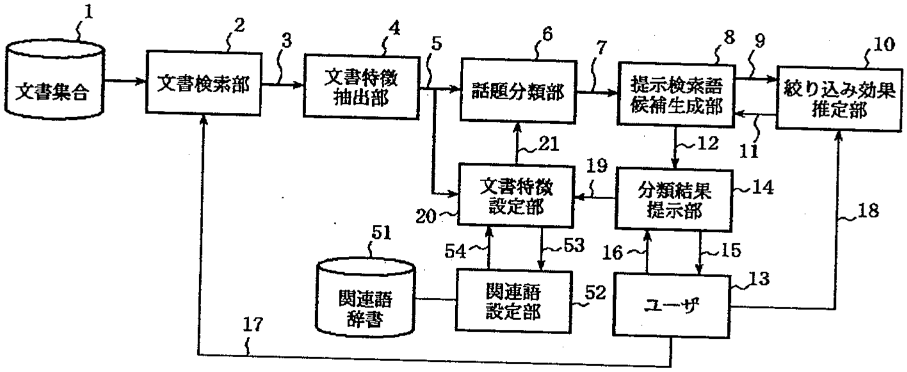


		検索語リスト			
		検索語1	検索語2	検索語3	・・・
文献識別子	文献1	3	0	0	・・・
	文献2	1	2	0	・・・
	文献3	0	0	5	・・・
	⋮	⋮	⋮	⋮	

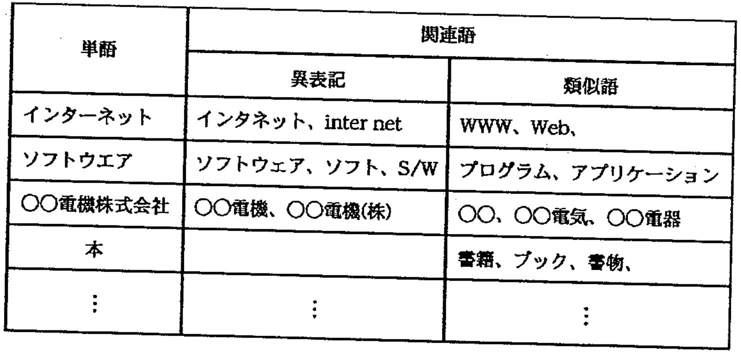
[Drawing 6]



[Drawing 7]

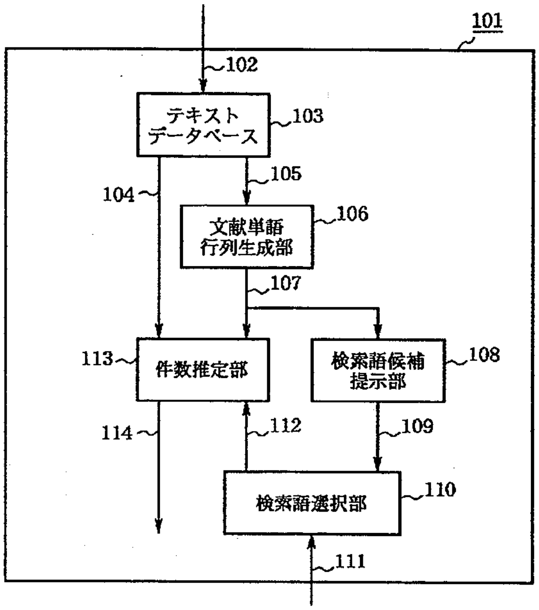


[Drawing 8]

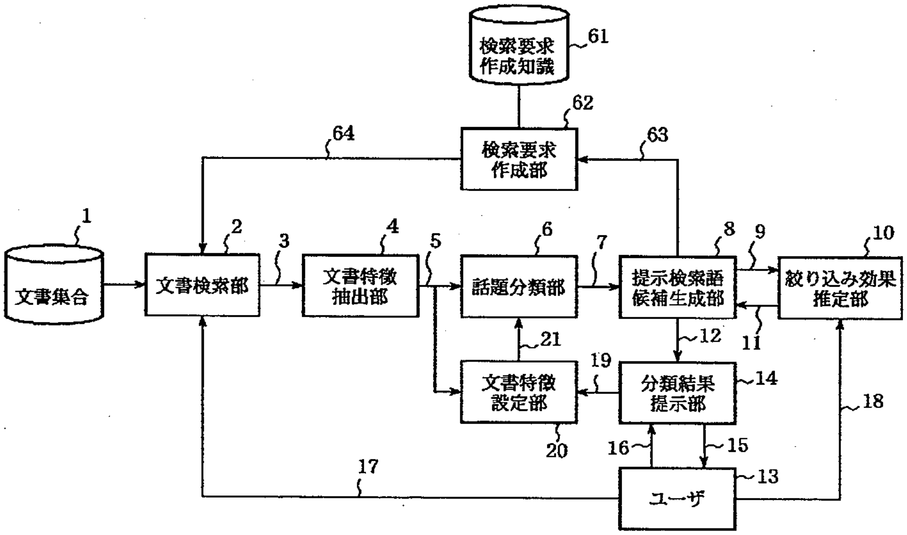


単語	関連語	
	異表記	類似語
インターネット	インタネット、inter net	WWW、Web、
ソフトウエア	ソフトウェア、ソフト、S/W	プログラム、アプリケーション
○○電機株式会社	○○電機、○○電機(株)	○○、○○電気、○○電器
本		書籍、ブック、書物、
⋮	⋮	⋮

[Drawing 10]



[Drawing 9]



[Translation done.]

(19)日本国特許庁（ＪＰ）　(12) **公 開 特 許 公 報**（Ａ）　(11)特許出願公開番号

特開2002－324077

（P2002－324077A）

(43)公開日　平成14年11月８日(2002.11.8)

(51)Int.Cl.^{7}	識別記号	ＦＩ	テーマコード（参考）
Ｇ０６Ｆ　17/30	３２０	Ｇ０６Ｆ　17/30	３２０Ｚ　５Ｂ０７５
	１７０		１７０Ａ
	２００		２００
	２１０		２１０Ｄ
	３４０		３４０Ｂ

審査請求　未請求　請求項の数20　ＯＬ　（全 16 頁）

(21)出願番号　特願2001－126541(P2001－126541)

(22)出願日　平成13年４月24日(2001.4.24)

(71)出願人　000006013
三菱電機株式会社
東京都千代田区丸の内二丁目２番３号

(72)発明者　永井　明人
東京都千代田区丸の内二丁目２番３号　三菱電機株式会社内

(72)発明者　高山　泰博
東京都千代田区丸の内二丁目２番３号　三菱電機株式会社内

(74)代理人　100066474
弁理士　田澤　博昭　（外１名）

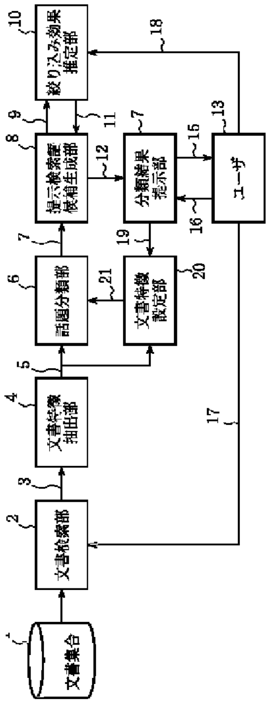
最終頁に続く

(54)【発明の名称】　文書検索装置および文書検索方法

(57)【要約】

【課題】　従来の文書検索装置は、効率的に絞り込み検索をするための検索語を選択することが難しい等の課題があった。

【解決手段】　文書を検索する文書検索部２と、文書ベクトル集合５を出力する文書特徴抽出部４と、文書ベクトル集合５を分類することによって話題を作成する話題分類部６と、絞り込み効果指標１１を算出する絞り込み効果推定部１０と、絞り込み効果指標１１が高い検索語を選択して提示検索語候補１２とし出力する提示検索語候補生成部８と、提示検索語候補１２と絞り込み効果指標１１とを話題別に提示する分類結果提示部１４と、文書ベクトル集合５を変更する文書特徴設定部２０とを備えるものである。



【特許請求の範囲】

【請求項１】　文書集合から検索条件に適合する文書を検索し出力する文書検索手段と、

前記文書に記述されている単語の出現頻度に基づいて、前記単語の統計的な重みを算出し、前記重みから得られる文書ベクトル集合を出力する文書特徴抽出手段と、

前記文書ベクトル集合を、文書ベクトル間の類似度に従って分類することによって話題を作成し、前記話題に属する文書情報と前記話題の重要度付き検索語とを出力する話題分類手段と、

前記文書情報と前記重要度付き検索語とを参照して前記話題に属する検索語の絞り込み効果指標を算出し出力する絞り込み効果推定手段と、

前記絞り込み効果指標を前記検索語に付与し、前記絞り込み効果指標が高い前記検索語を選択して提示検索語候補とし、該提示検索語候補と該提示検索語候補に対応する文書情報とを出力する提示検索語候補生成手段と、

前記提示検索語候補と前記話題とを対応付けて、前記絞り込み効果指標と共に提示し、指示情報または選択情報のどちらか一方もしくは両方を入力するように促す分類結果提示手段と、

前記選択情報に基づいて、前記文書ベクトル集合に含まれる文書ベクトルを変更して出力する文書特徴設定手段とを備える文書検索装置。

【請求項２】　絞り込み効果推定手段は、少なくとも一つまたは複数の話題に対して絞り込み効果指標を算出し出力することを特徴とする請求項１記載の文書検索装置。

【請求項３】　分類結果提示手段は、話題と該話題に属する提示検索語候補とを行列の形式で提示し、前記行列の各要素に絞り込み効果指標を提示することを特徴とする請求項１記載の文書検索装置。

【請求項４】　文書検索手段が出力する文書から該文書を特徴付ける単語を話題分類項目として抽出し、該話題分類項目に関連するテキストを参照して、前記話題分類項目に対する重みベクトルを学習し、前記話題分類項目と前記重みベクトルとを話題分類手段に出力する話題分類項目取得手段を備え、

前記話題分類手段は、前記話題分類項目と前記重みベクトルとに基づいて文書ベクトル集合を分類することによって話題を作成することを特徴とする請求項１記載の文書検索装置。

【請求項５】　話題分類項目取得手段は、文書検索手段が出力する文書から該文書に記述されている単語の出現頻度に基づいて話題分類項目を抽出することを特徴とする請求項４記載の文書検索装置。

【請求項６】　話題分類項目取得手段は、文書検索手段が出力する文書から該文書に記述されているタグを参照して話題分類項目を抽出することを特徴とする請求項４記載の文書検索装置。

【請求項７】　分類結果提示手段を介して指定された文書から文書ベクトルを算出し、文書特徴設定手段に出力する指定文書特徴抽出手段を備え、

前記文書特徴設定手段は、前記指定文書特徴抽出手段が出力する前記文書ベクトルと、文書特徴抽出手段が出力する文書ベクトル集合とに基づいて前記文書ベクトル集合を変更することを特徴とする請求項１記載の文書検索装置。

【請求項８】　所定の単語と関連する単語を定義し関連語として記録する第１の記録手段と、指定された検索語に対応する前記関連語を前記第１の記録手段から抽出して文書特徴設定手段に出力する関連語設定手段とを備え、

前記文書特徴設定手段は、前記関連語と分類結果提示手段から入力した選択情報とに基づいて、文書ベクトル集合を変更することを特徴とする請求項１記載の文書検索装置。

【請求項９】　検索要求文の作成知識を記録する第２の記録手段と、該第２の記録手段を参照して、提示検索語候補生成手段が出力した検索語に対応する検索要求文を作成し、文書検索手段に出力する検索要求作成手段とを備え、

前記提示検索語候補生成手段は、絞り込み効果指標に基づいて前記検索要求作成手段に出力する前記検索語を選択することを特徴とする請求項１記載の文書検索装置。

【請求項１０】　提示検索語候補生成手段は、複数の検索語を選択して検索要求作成手段に出力し、該検索要求作成手段は、複数の前記検索語に対する論理演算から検索要求文を作成することを特徴とする請求項９記載の文書検索装置。

【請求項１１】　文書集合から検索条件に適合する文書を検索し出力する文書検索ステップと、

前記文書に記述されている単語の出現頻度に基づいて、前記単語の統計的な重みを算出し、前記重みから得られる文書ベクトル集合を出力する文書特徴抽出ステップと、

前記文書ベクトル集合を、文書ベクトル間の類似度に従って分類することによって話題を作成し、前記話題に属する文書情報と前記話題の重要度付き検索語とを出力する話題分類ステップと、

前記文書情報と前記重要度付き検索語とを参照して前記話題に属する検索語の絞り込み効果指標を算出し出力する絞り込み効果推定ステップと、

前記絞り込み効果指標を前記検索語に付与し、前記絞り込み効果指標が高い前記検索語を選択して提示検索語候補とし、該提示検索語候補と該提示検索語候補に対応する文書情報とを出力する提示検索語候補生成ステップと、

前記提示検索語候補と前記話題とを対応付けて、前記絞り込み効果指標と共に提示し、指示情報または選択情報

のどちらか一方もしくは両方を入力するように促す分類結果提示ステップと、
前記選択情報に基づいて、前記文書ベクトル集合に含まれる文書ベクトルを変更して出力する文書特徴設定ステップとを有する文書検索方法。

【請求項１２】　絞り込み効果推定ステップは、少なくとも一つまたは複数の話題に対して絞り込み効果指標を算出し出力することを特徴とする請求項１１記載の文書検索方法。

【請求項１３】　分類結果提示ステップは、話題と該話題に属する提示検索語候補とを行列の形式で提示し、前記行列の各要素に絞り込み効果指標を提示することを特徴とする請求項１１記載の文書検索方法。

【請求項１４】　文書検索ステップが出力する文書から該文書を特徴付ける単語を話題分類項目として抽出し、該話題分類項目に関連するテキストを参照して、前記話題分類項目に対する重みベクトルを学習し、前記話題分類項目と前記重みベクトルとを出力する話題分類項目取得ステップを有し、
話題分類ステップは、前記話題分類項目と前記重みベクトルとに基づいて文書ベクトル集合を分類することによって話題を作成することを特徴とする請求項１１記載の文書検索方法。

【請求項１５】　話題分類項目取得ステップは、文書検索ステップが出力する文書から該文書に記述されている単語の出現頻度に基づいて話題分類項目を抽出することを特徴とする請求項１４記載の文書検索方法。

【請求項１６】　話題分類項目取得ステップは、文書検索ステップが出力する文書から該文書に記述されているタグを参照して話題分類項目を抽出することを特徴とする請求項１４記載の文書検索方法。

【請求項１７】　分類結果提示ステップを介して指定された文書から文書ベクトルを算出し出力する指定文書特徴抽出ステップを有し、
文書特徴設定ステップは、前記指定文書特徴抽出ステップが出力する前記文書ベクトルと、文書特徴抽出ステップが出力する文書ベクトル集合とに基づいて前記文書ベクトル集合を変更することを特徴とする請求項１１記載の文書検索方法。

【請求項１８】　所定の単語と関連する単語を定義し関連語として記録する第１の記録ステップと、指定された検索語に対応する前記関連語を抽出して出力する関連語設定ステップとを有し、
文書特徴設定ステップは、前記関連語と分類結果提示ステップから入力した選択情報とに基づいて、文書ベクトル集合を変更することを特徴とする請求項１１記載の文書検索方法。

【請求項１９】　検索要求文の作成知識を記録する第２の記録ステップと、提示検索語候補生成ステップが出力した検索語に対応する検索要求文を作成し出力する検索要求作成ステップとを有し、
前記提示検索語候補生成ステップは、絞り込み効果指標に基づいて出力する前記検索語を選択することを特徴とする請求項１１記載の文書検索方法。

【請求項２０】　提示検索語候補生成ステップは、複数の検索語を選択して出力し、検索要求作成ステップは、複数の前記検索語に対する論理演算から検索要求文を作成することを特徴とする請求項１９記載の文書検索方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、文書よりも大きな単位である話題毎に絞り込み効果を提示し、追加検索語の選択を容易にすることによって、効率よく文書の検索が実行できる文書検索装置および文書検索方法に関するものである。

【０００２】

【従来の技術】インターネットを利用して閲覧できるＨＴＭＬ文書や大規模なテキストデータベースに記録・管理される電子化文書などを検索した場合、検索結果として得られる文書情報の情報量が非常に多くなり、ユーザが所望する文書の発見に多くの時間と労力とが必要になってきている。このために、ユーザが目的の文書を効率的に検索できるように、検索結果として得られる文書情報を内容に応じて分類したり、絞り込み検索のために追加する検索語候補をユーザに提示したりするような、ユーザが所望する文書の発見を支援する技術に対する要求が高まっている。

【０００３】ユーザが目的の文書を効率的に検索できる技術に関しては、特開平９−２３１２３８号公報「テキスト検索結果表示方法及び装置」（以下、文献１と称する）、及び、特開平１１−２１３０００号公報「インタラクティブ情報検索方法及び装置及びインタラクティブ情報検索プログラムを格納した記憶媒体」（以下、文献２と称する）に開示された技術がある。

【０００４】文献１において開示された技術は、検索結果をファジィクラスタリングによって主題分類し、主題分類されたカテゴリを検索語の組と共にユーザに提示するものである。また、文献２において開示された技術は、検索結果をクラスタリングによって分類し、分類されたカテゴリを検索語の組と共にユーザに提示し、更に、ユーザが指定したカテゴリをサブカテゴリに分類して、インタラクティブな検索を可能にするものである。

【０００５】絞り込み検索を支援する技術に関しては、早川らの「ＷＷＷ検索サービスにおける検索結果絞り込み用インタフェースの開発」（情報処理学会，ヒューマンインタフェース研究会（ＨＩ７６−５），ｐｐ．２５，１９９８、以下、文献３と称する）において開示された技術がある。

【０００６】文献３において開示された技術は、検索語

と検索結果文献集合との関係を容易に把握できるようにするために、検索語が検索結果件数をどの程度絞り込めるかという情報を可視化してユーザに提示するインタフェースを提供している。また、検索語の絞り込み結果を、文献単語行列を用いて可視化しており、単語と文献とをそれぞれ行と列とに持つマトリクスを、直接ユーザに表の形で提示し、文献に対する単語の重みをマトリクスのセルの明るさで表現して、追加検索語の絞り込み効果を一覧できるようにしているものである。

【０００７】また、文献３において述べられた従来技術に関して、特開平１１－８５７６４号公報「検索結果件数の統計的推定方法及び装置及び検索結果件数の統計的推定プログラムを格納した記憶媒体」（以下、文献４と称する）に開示された技術がある。文献４において開示された技術は、文献３において検索結果をさらに絞り込むための検索語を追加した場合に、検索結果の件数がどのように変化するかを知るための検索結果件数の統計的推定方法について開示しているものである。

【０００８】図１０は、従来の文書検索装置を示す構成図であり、文献４に開示された実施の一例である推定装置の構成図である。図１０において、１０１は全文検索による文献検索を対象とする推定装置、１０２は文献を検索するための検索語、１０３は文献が記録・管理されているテキストデータベース、１０４はテキストデータベース１０３から出力される文献の検索結果総数、１０５は検索結果の中から５０件を無作為に抽出しテキストデータベース１０３から出力される文献サンプル集合、１０６は文献サンプル集合１０５が入力され各文献について各単語の出現回数を数える文献単語行列生成部、１０７は文献単語行列生成部１０６が生成した文献単語行列である。なお、文献サンプル集合１０５は、５０件に限定されるものではなく、状況に応じて適宜に設定できる。

【０００９】また、図１０において、１０８は文献単語行列１０７に対して単語重要度を計算する検索語候補提示部、１０９は検索語候補提示部１０８から出力される検索語候補、１１０は検索語候補１０９をモニタなどに出力する検索語選択部、１１１はユーザが検索語候補１０９を選択した際に入力される選択信号、１１２は選択信号１１１に基づいて検索語選択部１１０から出力される追加検索語、１１３は文献単語行列１０７と追加検索語１１２に基づいて追加検索語１１２の出現率を計算する件数推定部、１１４は件数推定部１１３から出力され検索結果総数１０４に出現率を乗じて得られる推定件数である。図１１は、従来の文書検索装置における文献単語行列１０７の一例を示す説明図である。

【００１０】次に動作について説明する。検索語１０２がテキストデータベース１０３に入力すると、検索語１０２に基づいてテキストデータベース１０３に記録・管理されている文献を検索する。テキストデータベース１０３は、検索結果の中から無作為に５０件を抽出し、文献サンプル集合１０５として文献単語行列生成部１０６に出力すると共に、検索結果総数１０４を件数推定部１１３に出力する。文献単語行列生成部１０６は、文献サンプル集合１０５に基づいて、各文献について各単語が出現する回数を数えることによって図１１に示された文献単語行列１０７を生成し、検索語候補提示部１０８及び件数推定部１１３に出力する。文献単語行列１０７において、行と列はそれぞれ文献識別子と検索語リストとを示し、表のセルの値は対応する行の文献の中に対応する列の検索語が出現する回数を示している。

【００１１】検索語候補提示部１０８は、文献単語行列１０７から助詞や代名詞などの単語を除き、任意の文献について任意の単語がどの程度重要であるかを示す単語重要度を計算する。単語重要度とは、特定の文献に集中的に出現している単語では高くなり、逆に多くの文献で出現している単語では低くなる指標である。また、検索語候補提示部１０８は、計算した単語重要度に基づいて、単語重要度が高い単語を検索語候補１０９として検索語選択部１１０に出力する。

【００１２】検索語選択部１１０は、検索語候補１０９をモニタ等に出力し、ユーザに検索語候補１０９から任意の検索語を選択させ、選択信号１１１として入力させる。また、検索語選択部１１０は、ユーザが入力した選択信号１１１に基づいて、追加検索語１１２として件数推定部１１３に出力する。

【００１３】件数推定部１１３には、検索結果総数１０４，文献単語行列１０７及び追加検索語１１２が入力する。件数推定部１１３は、文献単語行列１０７の追加検索語１１２の列が“０”ではない行の数を数え、これを全行数で除することにより追加検索語１１２の出現率が得られる。さらに、件数推定部１１３は、追加検索語１１２の出現率に検索結果総数１０４を乗じて得られる推定件数１１４を出力する。

【００１４】

【発明が解決しようとする課題】従来の文書検索装置は以上のように構成されているので、検索語の絞り込み効果に関して、文献１及び文献２に開示された従来技術では、検索結果の文書を分類して、分類されたカテゴリにおける検索語をユーザに提示しているが、提示された検索語を絞り込み検索に用いるために選択する際には、提示された検索語に対応する絞り込み効果の情報がないので、効率的に絞り込み検索をするための検索語を選択することが難しいという課題があった。

【００１５】また、従来の文書検索装置は、検索結果の可視化に関して、文献３に開示された従来技術では、検索語の絞り込み効果を文献単位でユーザに提示しているが、検索結果が数千程度の規模になると、文献単語行列による検索結果全体の絞り込み効果の表示が極めて困難になり、一覧性に欠けるという課題があった。

【００１６】また、従来の文書検索装置は、検索結果の可視化に関して、文献４に開示された従来技術では、検索結果から文献サンプル集合を出力し、文献サンプル集合に基づいて各文献毎に対応する追加検索語の出現率から、検索結果全体に対する絞り込み件数を推定しているので、推定された絞り込み件数をユーザに提示する際には、追加検索語毎の提示となるから、検索結果全体の絞り込み効果の表示が極めて困難になり、一覧性に欠けるという課題があった。

【００１７】また、従来の文書検索装置は、目的の文書への絞り込み検索に関して、文献２に開示された従来技術では、目的の文書に絞り込むために指定したカテゴリをサブカテゴリに詳細分類しているが、目的の文書は指定したカテゴリ内に全て存在しているわけではなく、他のカテゴリにもユーザが所望する文書が存在しており、絞り込み検索で得られる文書は指定したカテゴリに存在する文書のみに限定され、目的の文書の検索漏れが生じるから、絞り込み検索を再び行うことが困難であるという課題があった。

【００１８】また、従来の文書検索装置は、目的の文書への絞り込み検索に関して、文献３及び文献４に開示された従来技術では、初回の検索結果に対して検索語を追加していくＡＮＤ検索を行っているので、絞り込み検索の検索対象は、直前の検索語に対する検索結果の文書に限定され、全文書空間に散在する目的の文書の検索が不可能となり、絞り込み検索を再び行うことが困難であるという課題があった。

【００１９】この発明は上記のような課題を解決するためになされたもので、絞り込み検索のために提示する検索語に対して、絞り込み効果を表す指標を付与し、追加検索語の選択を容易にして、ユーザの絞り込み検索を効率的にする文書検索装置および文書検索方法を得ることを目的とする。

【００２０】また、この発明は、検索結果の文書を分類して、分類された文書集合を話題とし、文書よりも大きな単位である話題毎に検索語の絞り込み効果をユーザに提示して、検索結果の文書全体に対する絞り込み効果の一覧性を高めるようにする文書検索装置および文書検索方法を得ることを目的とする。

【００２１】さらに、この発明は、ユーザがフィードバック情報として指定した話題や検索語の情報を利用して、初回の検索結果を再分類することにより、話題間に分散した目的の文書を一つの話題に集めるようにして、検索結果の絞り込みを効率的にする文書検索装置および文書検索方法を得ることを目的とする。

【００２２】

【課題を解決するための手段】この発明に係る文書検索装置は、文書集合から検索条件に適合する文書を検索し出力する文書検索手段と、文書に記述されている単語の出現頻度に基づいて、単語の統計的な重みを算出し、重みから得られる文書ベクトル集合を出力する文書特徴抽出手段と、文書ベクトル集合を、文書ベクトル間の類似度に従って分類することによって話題を作成し、話題に属する文書情報と話題の重要度付き検索語とを出力する話題分類手段と、文書情報と重要度付き検索語とを参照して話題に属する検索語の絞り込み効果指標を算出し出力する絞り込み効果推定手段と、絞り込み効果指標を検索語に付与し、絞り込み効果指標が高い検索語を選択して提示検索語候補とし、当該提示検索語候補と当該提示検索語候補に対応する文書情報とを出力する提示検索語候補生成手段と、提示検索語候補と話題とを対応付けて、絞り込み効果指標と共に提示し、指示情報または選択情報のどちらか一方もしくは両方を入力するように促す分類結果提示手段と、選択情報に基づいて、文書ベクトル集合に含まれる文書ベクトルを変更して出力する文書特徴設定手段とを備えるものである。

【００２３】この発明に係る文書検索装置は、絞り込み効果推定手段が、少なくとも一つまたは複数の話題に対して絞り込み効果指標を算出し出力するようにしたものである。

【００２４】この発明に係る文書検索装置は、分類結果提示手段が、話題と当該話題に属する提示検索語候補とを行列の形式で提示し、行列の各要素に絞り込み効果指標を提示するようにしたものである。

【００２５】この発明に係る文書検索装置は、文書検索手段が出力する文書から当該文書を特徴付ける単語を話題分類項目として抽出し、話題分類項目に関連するテキストを参照して、話題分類項目に対する重みベクトルを学習し、話題分類項目と重みベクトルとを話題分類手段に出力する話題分類項目取得手段を備え、話題分類手段は、話題分類項目と重みベクトルとに基づいて文書ベクトル集合を分類することによって話題を作成するようにしたものである。

【００２６】この発明に係る文書検索装置は、話題分類項目取得手段が、文書検索手段から出力される文書から当該文書に記述されている単語の出現頻度に基づいて話題分類項目を抽出するようにしたものである。

【００２７】この発明に係る文書検索装置は、話題分類項目取得手段が、文書検索手段から出力される文書から当該文書に記述されているタグを参照して話題分類項目を抽出するようにしたものである。

【００２８】この発明に係る文書検索装置は、分類結果提示手段を介して指定された文書から文書ベクトルを算出し、文書特徴設定手段に出力する指定文書特徴抽出手段を備え、文書特徴設定手段が、指定文書特徴抽出手段から出力される文書ベクトルと、文書特徴抽出手段から出力される文書ベクトル集合とに基づいて文書ベクトル集合を変更するようにしたものである。

【００２９】この発明に係る文書検索装置は、所定の単語と関連する単語を定義し関連語として記録する第１の

記録手段と、指定された検索語に対応する関連語を第１の記録手段から抽出して文書特徴設定手段に出力する関連語設定手段とを備え、文書特徴設定手段が、関連語と分類結果提示手段から入力した選択情報とに基づいて、文書ベクトル集合を変更するようにしたものである。

【００３０】この発明に係る文書検索装置は、検索要求文の作成知識を記録する第２の記録手段と、当該第２の記録手段を参照して、提示検索語候補生成手段が出力した検索語に対応する検索要求文を作成し、文書検索手段に出力する検索要求作成手段とを備え、提示検索語候補生成手段が、絞り込み効果指標に基づいて検索要求作成手段に出力する検索語を選択するようにしたものである。

【００３１】この発明に係る文書検索装置は、提示検索語候補生成手段が、複数の検索語を選択して検索要求作成手段に出力し、検索要求作成手段が、複数の検索語に対する論理演算から検索要求文を作成するようにしたものである。

【００３２】この発明に係る文書検索方法は、文書集合から検索条件に適合する文書を検索し出力する文書検索ステップと、文書に記述されている単語の出現頻度に基づいて、単語の統計的な重みを算出し、重みから得られる文書ベクトル集合を出力する文書特徴抽出ステップと、文書ベクトル集合を、文書ベクトル間の類似度に従って分類することによって話題を作成し、話題に属する文書情報と話題の重要度付き検索語とを出力する話題分類ステップと、文書情報と重要度付き検索語とを参照して話題に属する検索語の絞り込み効果指標を算出し出力する絞り込み効果推定ステップと、絞り込み効果指標を検索語に付与し、絞り込み効果指標が高い検索語を選択して提示検索語候補とし、当該提示検索語候補と当該提示検索語候補に対応する文書情報とを出力する提示検索語候補生成ステップと、提示検索語候補と話題とを対応付けて、絞り込み効果指標と共に提示し、指示情報または選択情報のどちらか一方もしくは両方を入力するように促す分類結果提示ステップと、選択情報に基づいて、文書ベクトル集合に含まれる文書ベクトルを変更して出力する文書特徴設定ステップとを有するものである。

【００３３】この発明に係る文書検索方法は、絞り込み効果推定ステップが、少なくとも一つまたは複数の話題に対して絞り込み効果指標を算出し出力するようにしたものである。

【００３４】この発明に係る文書検索方法は、分類結果提示ステップが、話題と当該話題に属する提示検索語候補とを行列の形式で提示し、行列の各要素に絞り込み効果指標を提示するようにしたものである。

【００３５】この発明に係る文書検索方法は、文書検索ステップが出力する文書から当該文書を特徴付ける単語を話題分類項目として抽出し、話題分類項目に関連するテキストを参照して、話題分類項目に対する重みベクトルを学習し、話題分類項目と重みベクトルとを出力する話題分類項目取得ステップを有し、話題分類ステップが、話題分類項目と重みベクトルとに基づいて文書ベクトル集合を分類することによって話題を作成するようにしたものである。

【００３６】この発明に係る文書検索方法は、話題分類項目取得ステップが、文書検索ステップから出力される文書から当該文書に記述されている単語の出現頻度に基づいて話題分類項目を抽出するようにしたものである。

【００３７】この発明に係る文書検索方法は、話題分類項目取得ステップが、文書検索ステップから出力される文書から当該文書に記述されているタグを参照して話題分類項目を抽出するようにしたものである。

【００３８】この発明に係る文書検索方法は、分類結果提示ステップを介して指定された文書から文書ベクトルを算出し出力する指定文書特徴抽出ステップを有し、文書特徴設定ステップが、指定文書特徴抽出ステップから出力された文書ベクトルと、文書特徴抽出ステップから出力された文書ベクトル集合とに基づいて文書ベクトル集合を変更するようにしたものである。

【００３９】この発明に係る文書検索方法は、所定の単語と関連する単語を定義し関連語として記録する第１の記録ステップと、指定された検索語に対応する関連語を抽出して出力する関連語設定ステップとを有し、文書特徴設定ステップが、関連語と分類結果提示ステップから入力した選択情報とに基づいて、文書ベクトル集合を変更するようにしたものである。

【００４０】この発明に係る文書検索方法は、検索要求文の作成知識を記録する第２の記録ステップと、提示検索語候補生成ステップが出力した検索語に対応する検索要求文を作成し出力する検索要求作成ステップとを有し、提示検索語候補生成ステップが、絞り込み効果指標に基づいて出力する検索語を選択するようにしたものである。

【００４１】この発明に係る文書検索方法は、提示検索語候補生成ステップが、複数の検索語を選択して出力し、検索要求作成ステップが、複数の検索語に対する論理演算から検索要求文を作成するようにしたものである。

【００４２】

【発明の実施の形態】以下、この発明の実施の一形態を説明する。

実施の形態１．図１は、この発明の実施の形態１による文書検索装置の構成を示すブロック図である。図１において、１は検索対象となる文書集合であり、例えばインターネットを利用して閲覧できるＨＴＭＬ文書や、インターネットを利用して送受信できる電子メール、記録装置や記録媒体に記録された大規模なテキストデータベースに記録・管理される電子化文書などの電子化されたテキストから構成される。２は所定の条件に従って文書集

合１から文書を検索する文書検索部（文書検索手段）、３は文書検索部２が所定の条件に従って文書集合１から検索した結果である検索結果文書集合、４は文書検索部２が出力した検索結果文書集合３に対応する文書ベクトルを作成する文書特徴抽出部（文書特徴抽出手段）である。なお、文書ベクトルとは文書毎の各単語の重みをベクトルの形式で表現したものである。

【００４３】また、図１において、５は文書特徴抽出部４が作成する文書ベクトルに基づいて出力される文書ベクトル集合、６は文書特徴抽出部４が出力した文書ベクトル集合５に基づいて算出した文書ベクトル間の類似度に従って文書ベクトル集合５を複数の集合に分類することで話題を作成する話題分類部（話題分類手段）、７は話題分類部６で分類された各話題の文書情報と共に出力される重要度付き検索語集合、８は話題分類部６が出力した重要度付き検索語集合７から所定の基準で提示検索語候補を選択する提示検索語候補生成部（提示検索語候補生成手段）である。提示検索語候補を選択するための所定の基準とは、例えば重要度の順に上位一定数を選択する。

【００４４】さらに、図１において、９は提示検索語候補生成部８から各話題の文書情報と共に出力される各話題の重要度付き検索語集合、１０は重要度付き検索語集合９からユーザが指定した話題に属する検索語の絞り込み効果を推定する絞り込み効果推定部（絞り込み効果推定手段）、１１は絞り込み効果を推定するための指標となる絞り込み効果推定部１０が算出した絞り込み効果指標、１２は絞り込み効果指標１１に基づいて提示検索語候補生成部８が選択し対応する話題に属する文書情報や絞り込み効果指標１１と共に出力する提示検索語候補、１３は文書検索装置を操作するユーザ、１４は提示検索語候補１２を各話題に対応付けて絞り込み効果指標１１と共にユーザ１３に提示する分類結果提示部（分類結果提示手段）、１５は分類結果提示部１４がユーザ１３に提示するために視覚化した分類結果、１６はユーザ１３が分類結果提示部１４に送信する指示情報である。

【００４５】さらに、図１において、１７はユーザ１３が文書検索部２に入力する検索条件、１８はユーザ１３が絞り込み効果推定部１０に入力するユーザ１３により指定された話題、１９はユーザ１３が検索結果の再分類を指示した場合に分類結果提示部１４から出力されるユーザ１３の選択情報、２０はユーザ１３の選択情報１９に基づいて文書ベクトルや検索語に対する重みを変更する文書特徴設定部（文書特徴設定手段）、２１は文書特徴設定部２０が話題分類部６に出力する変更された文書ベクトルである。

【００４６】図２は、この発明の実施の形態１による文書検索装置の動作を説明するフローチャートである。図３は、この発明の実施の形態１における文書ベクトルの一例を示す説明図である。図４は、この発明の実施の形態１における検索語－話題対応表の一例を示す説明図である。

【００４７】次に動作について説明する。先ず、ステップＳＴ１において、ユーザ１３は文書検索部２に検索条件１７を入力する。検索条件１７は、例えば検索語や複数の検索語同士の論理的な式である。次に、ステップＳＴ２において、文書検索部２は、入力された検索条件１７に基づいて文書集合１の文書を検索し、検索条件１７に適合する検索結果文書集合３を出力する（文書検索ステップ）。検索対象となる文書集合１は、例えばインターネットを利用して閲覧できるＨＴＭＬ文書や、インターネットを利用して送受信できる電子メール、記録装置や記録媒体に記録された大規模なテキストデータベースに記録・管理される電子化文書などの電子化されたテキストである。また、文書検索部２は、検索条件１７で検索可能であればよく、例えばインターネットにおいて一般的に使用されている全文検索エンジン等を用いてもよい。さらに、文書検索部２は、検索結果の文書に関する種々の情報、例えば文書を特定するための文書ＩＤ番号や文書ファイルの場所、文書のタイトル等の情報を文書情報として検索結果文書集合３と共に出力する。

【００４８】次に、ステップＳＴ３において、検索結果文書集合３の各文書に対する文書ベクトルを文書特徴抽出部４が求める。文書ベクトルは、図３に示されたように、文書毎の各単語の出現頻度に基づいて、各文書Ｄ１，Ｄ２，・・・，Ｄｊ，・・・Ｄｍ、に対する各単語ＫＷ１，ＫＷ２，・・・，ＫＷｉ，・・・，ＫＷｎの統計的な重みＷｉｊを算出し、文書毎の各単語の重みをベクトルの形式で表現したものである。この統計的な重みの算出方法は、ＴＦ・ＩＤＦやχ２乗統計値など種々の算出方法があり、目的に合わせて適宜に選択して用いればよい。文書特徴抽出部４は、統計的な重みを算出して得られた文書ベクトル集合５を出力する（文書特徴抽出ステップ）。

【００４９】次に、ステップＳＴ４において、話題分類部６は、文書ベクトル間の類似度を算出し、類似度に従って文書ベクトル集合５を複数の集合に分類することで話題を作成する（話題分類ステップ）。文書を分類する方法としては、トップダウンに分類カテゴリを与えて分類する文書分類と、ボトムアップに類似する文書をまとめあげていくクラスタリングとの２種類に大別される。

【００５０】文書分類では、分類先のカテゴリを予め設定して、カテゴリに属するサンプル文書から、分類カテゴリに対する統計的な重みを分類カテゴリベクトルとして学習しておき、入力された文書ベクトル集合５の各文書ベクトルと、分類カテゴリベクトルとの類似度を算出する。類似度は、例えばベクトルの内積値を用いる。このようにして得られた類似度に基づいて、最も類似度が高い分類カテゴリに文書を分類する。一方、クラスタリングは、入力された文書ベクトル集合５に存在する全て

の文書ベクトル間の類似度を算出し、類似度が高い文書ベクトル同士をまとめて一つのクラスタとし、クラスタに対する類似度の算出とまとめあげの処理とを繰り返すことによって文書を分類する。

【００５１】話題分類部６は、文書ベクトルを分類する機能があればよく、上述した文書分類とクラスタリングに限られるものではなく、その他の分類方法（例えば主成分分析）を採用してもよい。また、話題分類部６は、分類された集合を話題として、各話題毎に算出した重要度の高い単語を検索語とする。例えば、ある話題に属する全文書中の単語について、不要とみなして別途設定した単語を削除した上で、各単語の出現頻度を数えて重要度とし、重要度の上位一定数を検索語とする。さらに、話題分類部６は、各話題に属する文書情報と、各話題の重要度付き検索語集合７とを出力する。

【００５２】次に、ステップＳＴ５において、ユーザ１３の検索が初回である場合はステップＳＴ６に進み、ユーザ１３の検索が初回ではない場合はステップＳＴ７に進む。ステップＳＴ６において、提示検索語候補生成部８は、各話題の重要度付き検索語集合７から所定の基準で提示検索語候補１２を選択し、ステップＳＴ８に進む。提示検索語候補１２を選択するための所定の基準とは、例えば重要度の順に上位一定数を選択する。一方、ステップＳＴ７において、絞り込み効果推定部１０によって算出された絞り込み効果指標１１を各検索語候補に付与して、絞り込み効果指標１１が高い検索語候補を選択して提示検索語候補１２とする。提示検索語候補生成部８は、このようにして選択された提示検索語候補１２を、対応する話題に属する文書情報と共に分類結果提示部１４に出力する（提示検索語候補生成ステップ）。

【００５３】次に、ステップＳＴ８において、分類結果提示部１４は、提示検索語候補１２を、各話題に対応付けて絞り込み効果指標１１と共に視覚化した分類結果１５としてユーザ１３に提示する（分類結果提示ステップ）。視覚化の方法は、例えば図４に示されたように、検索語−話題対応表を用いる。検索語−話題対応表は、提示検索語候補１２と対応する話題とが行列の形式によって表現されており、行列の各要素には視覚化された情報として絞り込み効果指標１１をユーザ１３に提示する。また、分類結果提示部１４は、対応する話題に属する文書情報を用いて、各話題Ｔ１，Ｔ２，・・・，Ｔ６の何れかをユーザ１３が指定すると、指定された話題に属する文書の一覧、及び各種の文書情報が参照できるようにする。

【００５４】次に、ステップＳＴ９において、ユーザ１３が分類結果提示部１４に提示された検索語−話題対応表を参照して、話題を指定する場合にはステップＳＴ１０に進み、ユーザ１３が話題を指定しない場合にはステップＳＴ１２に進む。ステップＳＴ１０において、ユーザ１３は、指定する話題１８を絞り込み効果推定部１０に入力する。次に、ステップＳＴ１１において、各話題に属する文書情報及び重要度付き検索語集合９を、提示検索語候補生成部８を参照して入手し、ユーザ１３が指定した話題１８に属する検索語の絞り込み効果指標１１を算出する（絞り込み効果推定ステップ）。絞り込み効果指標１１の算出方法としては、例えばユーザ１３が指定した話題１８に属する検索語に関して、各話題との統計的な重みを再び算出する。または、ユーザ１３が指定した話題１８に属する検索語を含む文書数を絞り込み効果指標１１とする。このようにして算出された絞り込み効果指標１１を提示検索語候補生成部８に出力し、さらに、提示検索語候補生成部８から分類結果提示部１４に出力され、分類結果提示部１４が図４に示されたような検索語−話題対応表のように視覚化し、ユーザ１３が指定した話題１８に属する提示検索語候補１２と各話題との対応で絞り込み効果指標１１をユーザ１３に提示する。ステップＳＴ１１の処理が終了した場合には、ステップＳＴ７に戻る。

【００５５】なお、絞り込み効果推定部１０に入力する話題１８に関して、複数の話題を指定できるようにした場合は、話題の選択幅が広がるので、ユーザ１３が所望する文書の内容を、より的確に指定できるようになる。例えば、複数指定された話題に属する検索語によって重要度の高い順に上位一定個数を選択し、選択された提示検索語候補１２と各話題との対応によって絞り込み効果指標１１をユーザ１３に提示すればよい。

【００５６】また、絞り込み効果推定部１０が算出する絞り込み効果指標１１を、検索語単位ではなく、話題単位で算出することによって、より広い範囲の絞り込み効果指標１１の傾向が検索結果全体に対して把握できるようになる。例えば、指定された話題に属する文書中の単語出現頻度に基づいて、話題の重みベクトルを作成し、各話題との類似度をベクトルの内積値として求めて、指定された話題と各話題との類似度を行列の形式によりユーザ１３に提示すればよい。

【００５７】一方、ステップＳＴ１２において、ユーザ１３が分類結果提示部１４が提示した検索語−話題対応表を参照して、絞り込み効果推定部１０に対して話題１８を入力せずに、再分類を指示した場合にはステップＳＴ１３に進み、再分類を指示しない場合にはこの処理を終了する。

【００５８】ステップＳＴ１３において、ユーザ１３は分類結果提示部１４に対して、探したい文書に近い内容の話題や検索語を指定するための指示情報１６を入力し、再分類の指示を与える。分類結果提示部１４は、指定された話題に属する文書の文書ＩＤ番号や、指定された検索語などの情報をユーザ１３の選択情報１９として、文書特徴設定部２０に出力する。次に、ステップＳＴ１４において、文書特徴設定部２０は、ユーザ１３の選択情報１９に基づいて、入力された文書ＩＤ番号に対

応する文書ベクトルや指定された検索語に対する重みを変更して、変更された文書ベクトル２１を話題分類部６に出力する（文書特徴設定ステップ）。重みの変更は、例えば予め設定した定数値を重みに加算することによって、指定された話題に属する文書や検索語の重みを変更する。ステップＳＴ１４の処理が終了した場合には、ステップＳＴ４に戻る。

【００５９】以上のように、この実施の形態１によれば、ユーザ１３が指定した話題１８に属する検索語の絞り込み効果指標１１を算出する絞り込み効果推定部１０と、絞り込み効果指標１１を視覚化してユーザ１３に提示する分類結果提示部１４とを備え、ユーザ１３が指定した話題１８に属する提示検索語候補１２と各話題とに対応する絞り込み効果指標１１をユーザ１３に提示するようにしたので、ユーザ１３が追加検索語の選択を容易にできるから、絞り込み効果の高い追加検索語を的確に選択できると共に、絞り込み検索の効率がよくなるという効果が得られる。

【００６０】また、この実施の形態１によれば、文書ベクトル集合５に基づいて算出した文書ベクトル間の類似度に従って文書ベクトル集合５を複数の集合に分類することで話題を作成する話題分類部６と、話題毎に検索語の絞り込み効果指標１１を検索語－話題対応表のように視覚化しユーザ１３に提示する分類結果提示部１４とを備え、文書単位よりも大きな単位である話題単位で絞り込み効果を提示するようにしたから、検索結果の文書全体に対する絞り込み効果の一覧性が高まり、検索結果全体の内容を効率的に把握できるという効果が得られる。

【００６１】また、この実施の形態１によれば、ユーザ１３がフィードバック情報として指示する指示情報１６に基づいて文書ベクトルの変更を行う文書特徴設定部２０を備え、初回の検索結果に対して再分類するようにしたので、話題間に分散した目的の文書を一つの話題に集めるようにしたから、目的の文書への絞り込み検索の効率がよくなるという効果が得られる。

【００６２】実施の形態２．図５は、この発明の実施の形態２による文書検索装置の構成を示すブロック図である。図５において、図１と同一符号は同一または相当部分を示すのでその説明を省略する。３１は検索結果文書集合３における文書から、文書を特徴付ける単語を話題分類項目として抽出し、抽出された話題分類項目に関連するテキストを参照して、話題分類項目に対する重みベクトルを学習する話題分類項目取得部（話題分類項目取得手段）である。３２は話題分類項目取得部３１が話題分類部６に出力し、話題分類項目と重みベクトルとを含む話題分類項目情報である。

【００６３】次に動作について説明する。実施の形態２において、文書集合１、文書検索部２、文書特徴抽出部４、提示検索語候補生成部８、絞り込み効果推定部１０、ユーザ１３、分類結果提示部１４、文書特徴設定部２０等の動作、及びこれらが奏する効果については、実施の形態１と同様であるのでその説明を省略する。

【００６４】話題分類項目取得部３１は、検索結果文書集合３における文書から、文書を特徴付ける単語を話題分類項目として抽出し、抽出された話題分類項目に関連するテキストを参照して、話題分類項目に対する重みベクトルを学習し、話題分類項目を重みベクトルと共に話題分類項目情報３２として話題分類部６に出力する（話題分類項目取得ステップ）。

【００６５】話題分類項目としては、例えば文書中において出現率が高い単語や、ＨＴＭＬ文書に記述されているＨＴＭＬタグを参照して得られるＨＴＭＬ文書のタイトルやＵＲＬに含まれるドメイン名などである。抽出された話題分類項目に関連するテキストは、例えば、話題分類項目が存在する位置周辺のテキストを解析し、段落区切り，章立て，箇条書き，リンク先などの特定のＨＴＭＬタグを検出し、検出されたＨＴＭＬタグに関連付けられたテキストを複写することにより、話題分類項目に関連するテキストを抽出する。

【００６６】また、話題分類項目取得部３１は、抽出された話題分類項目に関連するテキストを用いて、話題分類項目に対する重みベクトルを学習し、話題分類項目を重みベクトルと共に話題分類部６に出力する。話題分類部６は、話題分類項目取得部３１から入力した話題分類項目情報３２と、文書特徴抽出部４から入力した文書ベクトル集合５の各文書ベクトルとの類似度を、例えばベクトルの内積値により算出することによって、最も類似度が高い話題分類項目に文書を分類する。

【００６７】以上のように、この実施の形態２によれば、実施の形態１と同様の効果を奏すると共に、検索結果文書集合３における文書から話題分類項目を抽出し、当該話題分類項目に対する重みベクトルと共に出力する話題分類項目取得部３１を備え、話題分類項目情報３２と各文書ベクトルとの類似度に基づいて文書を分類するようにしたので、話題分類部６で用いる話題分類項目が自動的に取得でき、予め分類カテゴリを設定しておく必要がなくなるから、検索結果の文書項目の設定作業が不要になると共に、ユーザ１３の絞り込み検索の効率がよくなるという効果が得られる。

【００６８】実施の形態３．図６は、この発明の実施の形態３による文書検索装置の構成を示すブロック図である。図６において、図１と同一符号は同一または相当部分を示すのでその説明を省略する。４１はユーザ１３から指定された文書から文書ベクトルを算出する指定文書特徴抽出部（指定文書特徴抽出手段）、４２は指定文書特徴抽出部４１が文書特徴設定部２０に出力する文書ベクトル、４３はユーザ１３が指定文書特徴抽出部４１に出力する指定された文書である。

【００６９】次に動作について説明する。実施の形態３において、文書集合１、文書検索部２、文書特徴抽出部

４、話題分類部６、提示検索語候補生成部８、絞り込み効果推定部１０、分類結果提示部１４等の動作、及びこれらが奏する効果については、実施の形態１と同様であるのでその説明を省略する。

【００７０】ユーザ１３は、分類結果提示部１４を参照して、ユーザ１３が所望する文書に近い内容の文書を選択し、指定文書特徴抽出部４１に指定する文書４３を指示する。指定文書特徴抽出部４１は、指定された文書４３に含まれる単語の出現回数に基づいて統計的な重みを計算して文書ベクトル４２を算出し、文書特徴設定部２０に出力する。文書特徴設定部２０は、指定された文書４３の文書ベクトル４２と、文書特徴抽出部４から入力した文書ベクトル集合５との類似度を計算し、文書ベクトル集合５における文書ベクトルの重みを変更する（指定文書特徴抽出ステップ）。例えば、類似度の高い順に上位一定個数の文書ベクトルを文書ベクトル集合５から選択し、類似度を文書ベクトルの重みに加算して変更する。話題分類部６は、変更された文書ベクトル集合５に対して分類を行う。

【００７１】以上のように、この実施の形態３によれば、実施の形態１と同様の効果を奏すると共に、ユーザ１３から指定された文書から文書ベクトルを算出する指定文書特徴抽出部４１を備え、指定文書特徴抽出部４１が出力する文書ベクトル４２と文書特徴抽出部４が出力した文書ベクトル集合５との類似度を計算し、文書ベクトルの重みを変更するようにしたので、ユーザ１３が所望する文書に近い内容の文書４３を直接指定することによって、文書特徴設定部２０の文書ベクトルの変更が可能になると共に、ユーザ１３の絞り込み検索の効率がよくなるという効果が得られる。

【００７２】実施の形態４．図７は、この発明の実施の形態４による文書検索装置の構成を示すブロック図である。図７において、図１と同一符号は同一または相当部分を示すのでその説明を省略する。５１は単語と当該単語に関連する関連語が定義されている関連語辞書（第１の記録手段）、５２は検索語が入力されると検索語に関連する関連語を関連語辞書５１から選択し関連語を出力する関連語設定部（関連語設定手段）、５３は文書特徴設定部２０が出力し関連語設定部５２に入力する検索語、５４は関連語設定部５２が出力し文書特徴設定部２０に入力する関連語である。

【００７３】図８は、この発明の実施の形態４における単語と関連語とを定義した一例を示す説明図である。図８において、関連語は、異表記と類似語とを関連語としており、それぞれの行に記述される単語と対応している。

【００７４】次に動作について説明する。実施の形態４において、文書集合１、文書検索部２、文書特徴抽出部４、話題分類部６、提示検索語候補生成部８、絞り込み効果推定部１０、ユーザ１３、分類結果提示部１４等の動作、及びこれらが奏する効果については、実施の形態１と同様であるのでその説明を省略する。

【００７５】関連語辞書５１には、図８に示されたように、所定の単語と当該単語に関連する単語とを予め記録しておく（第１の記録ステップ）。関連語設定部５２は、関連語辞書５１を参照し、入力した検索語５３に対応する異表記と類似語とを抽出して関連語５４とし、文書特徴設定部２０に関連語５４を出力する（関連語設定ステップ）。文書特徴設定部２０は、関連語設定部５２から入力した関連語５４を、分類結果提示部１４から入力したユーザ１３の選択情報１９に追加し、文書特徴抽出部４から入力した文書ベクトル集合５を変更する。

【００７６】以上のように、この実施の形態４によれば、実施の形態１と同様の効果を奏すると共に、関連語辞書５１を参照して関連語５４を出力する関連語設定部５２を備え、関連語設定部５２から入力する関連語５４を、分類結果提示部１４から入力したユーザ１３の選択情報１９に追加し、文書ベクトル集合５を変更するようにしたので、検索語５３の異表記や類似語などを含む文書が検索できるようになるから、検索漏れが抑制されると共に、ユーザ１３が所望する文書を発見するための検索の効率がよくなるという効果が得られる。

【００７７】実施の形態５．図９は、この発明の実施の形態５による文書検索装置の構成を示すブロック図である。図９において、図１と同一符号は同一または相当部分を示すのでその説明を省略する。６１は磁気記録装置などで構成され検索要求の作成知識を記録する検索要求作成知識（第２の記録手段）、６２は検索要求検索語が入力されると検索要求検索語に基づく最適な検索要求文を検索要求作成知識６１から選択し検索要求文を出力する検索要求作成部（検索要求作成手段）、６３は提示検索語候補生成部８が出力し検索要求作成部６２に入力する検索要求検索語、６４は検索要求作成部６２が出力し文書検索部２に入力する検索要求文である。

【００７８】次に動作について説明する。実施の形態５において、文書集合１、文書特徴抽出部４、話題分類部６、絞り込み効果推定部１０、ユーザ１３、分類結果提示部１４、文書特徴設定部２０等の動作、及びこれらが奏する効果については、実施の形態１と同様であるのでその説明を省略する。

【００７９】提示検索語候補生成部８は、絞り込み効果推定部１０から入力された絞り込み効果指標１１を参照し、絞り込み効果指標１１の高い検索語を検索要求作成部６２に検索要求検索語６３として出力する。検索要求検索語６３は複数であってもよい。検索要求作成知識６１には、文書検索部２に対して検索処理の実行を指示する検索要求文６４を作成するための知識が予め定義されている（第２の記録ステップ）。例えば、検索要求文６４は、検索命令と検索条件との２種類から構成される。検索命令としては、例えば、＜検索実行＞，＜実行状態

取得＞，＜データベース指定＞などのコマンドの種類を定義する。また、検索条件としては、検索語，検索語間の論理演算子，検索結果として得る情報の指定などのパラメータの記述形式を定義する。

【００８０】検索要求作成部６２は、検索要求作成知識６１を参照して、検索要求文６４の定義に従って検索要求検索語６３を検索条件に設定する（検索要求作成ステップ）。検索要求文６４の検索命令は、例えば＜検索実行＞とし、検索要求文６４を作成して文書検索部２に出力する。また、検索要求文６４の検索条件を設定する際には、複数の検索要求検索語６３に対して例えばＡＮＤ演算子で記述し、検索要求検索語６３に付与された絞り込み効果指標１１が予め設定した閾値以上であれば、絞り込み効果が高いとみなして、より広範囲の文書を検索するように、ＯＲ演算子で記述することもできる。

【００８１】以上のように、この実施の形態５によれば、実施の形態１と同様の効果を奏すると共に、検索要求作成知識６１を参照して検索要求文６４を作成する検索要求作成部６２を備え、検索要求作成部６２が出力する検索要求文６４に基づいて再び検索するようにしたので、初回の検索結果に含まれなかったユーザ１３が所望する文書を文書集合１から自動的に検索できるようになると共に、ユーザ１３の絞り込み検索の効率がよくなるという効果が得られる。

【００８２】

【発明の効果】以上のように、この発明によれば、文書集合から検索条件に適合する文書を検索し出力する文書検索手段と、文書に記述されている単語の出現頻度に基づいて、単語の統計的な重みを算出し、重みから得られる文書ベクトル集合を出力する文書特徴抽出手段と、文書ベクトル集合を、文書ベクトル間の類似度に従って分類することによって話題を作成し、話題に属する文書情報と話題の重要度付き検索語とを出力する話題分類手段と、文書情報と重要度付き検索語とを参照して話題に属する検索語の絞り込み効果指標を算出し出力する絞り込み効果推定手段と、絞り込み効果指標を検索語に付与し、絞り込み効果指標が高い検索語を選択して提示検索語候補とし、当該提示検索語候補と当該提示検索語候補に対応する文書情報とを出力する提示検索語候補生成手段と、提示検索語候補と話題とを対応付けて、絞り込み効果指標と共に提示し、指示情報または選択情報のどちらか一方もしくは両方を入力するように促す分類結果提示手段と、選択情報に基づいて、文書ベクトル集合に含まれる文書ベクトルを変更して出力する文書特徴設定手段とを備えるように構成したので、絞り込み検索のために提示する提示検索語候補に対して絞り込み効果指標を付与したから、追加検索語の選択を容易にすることができ、絞り込み検索を効率よく実行できるという効果を奏する。

【００８３】この発明によれば、絞り込み効果推定手段が、少なくとも一つまたは複数の話題に対して絞り込み効果指標を算出し出力するように構成したので、文書単位よりも大きな単位である話題単位で絞り込み検索を実行できるから、より広い範囲の検索結果に対する絞り込み効果指標の傾向が把握できると共に、ユーザが所望する文書の内容を的確に指定できるという効果を奏する。

【００８４】この発明によれば、分類結果提示手段が、話題と当該話題に属する提示検索語候補とを行列の形式で提示し、行列の各要素に絞り込み効果指標を提示するように構成したので、検索結果の文書全体に対する絞り込み効果の一覧性が高まり、検索結果全体の内容を効率的に把握できるという効果を奏する。

【００８５】この発明によれば、文書検索手段が出力する文書から当該文書を特徴付ける単語を話題分類項目として抽出し、話題分類項目に関連するテキストを参照して、話題分類項目に対する重みベクトルを学習し、話題分類項目と重みベクトルとを話題分類手段に出力する話題分類項目取得手段を備え、話題分類手段は、話題分類項目と重みベクトルとに基づいて文書ベクトル集合を分類することによって話題を作成するように構成したので、話題分類手段で用いる話題分類項目が自動的に取得でき、予め分類カテゴリを設定しておく必要がなくなるから、検索結果の文書項目の設定作業が不要になると共に、絞り込み検索の効率がよくなるという効果を奏する。

【００８６】この発明によれば、話題分類項目取得手段が、文書検索手段から出力される文書から当該文書に記述されている単語の出現頻度に基づいて話題分類項目を抽出するように構成したので、話題分類項目を自動的に効率よく抽出することができるという効果を奏する。

【００８７】この発明によれば、話題分類項目取得手段が、文書検索手段から出力される文書から当該文書に記述されているタグを参照して話題分類項目を抽出するように構成したので、タグが記述されている文書から話題分類項目を自動的に効率よく抽出することができるという効果を奏する。

【００８８】この発明によれば、分類結果提示手段を介して指定された文書から文書ベクトルを算出し、文書特徴設定手段に出力する指定文書特徴抽出手段を備え、文書特徴設定手段が、指定文書特徴抽出手段から出力される文書ベクトルと、文書特徴抽出手段から出力される文書ベクトル集合とに基づいて文書ベクトル集合を変更するように構成したので、ユーザが所望する文書に近い内容の文書を直接指定することによって、文書特徴設定手段の文書ベクトルの変更が可能になると共に、絞り込み検索の効率がよくなるという効果を奏する。

【００８９】この発明によれば、所定の単語と関連する単語を定義し関連語として記録する第１の記録手段と、指定された検索語に対応する関連語を第１の記録手段から抽出して文書特徴設定手段に出力する関連語設定手段

とを備え、文書特徴設定手段が、関連語と分類結果提示手段から入力した選択情報とに基づいて、文書ベクトル集合を変更するように構成したので、検索語の異表記や類似語などを含む文書が検索できるようになるから、検索漏れが抑制されると共に、ユーザが所望する文書を発見するための検索の効率がよくなるという効果を奏する。

【００９０】この発明によれば、検索要求文の作成知識を記録する第２の記録手段と、当該第２の記録手段を参照して、提示検索語候補生成手段が出力した検索語に対応する検索要求文を作成し、文書検索手段に出力する検索要求作成手段とを備え、提示検索語候補生成手段が、絞り込み効果指標に基づいて検索要求作成手段に出力する検索語を選択するように構成したので、初回の検索結果に含まれなかったユーザが所望する文書を文書集合から自動的に検索できるようになると共に、ユーザの絞り込み検索の効率がよくなるという効果を奏する。

【００９１】この発明によれば、提示検索語候補生成手段が、複数の検索語を選択して検索要求作成手段に出力し、検索要求作成手段が、複数の検索語に対する論理演算から検索要求文を作成するように構成したので、ＡＮＤ演算ではより的確に絞り込み検索が実行でき、ＯＲ演算ではより広範囲に絞り込み検索が実行できるという効果を奏する。

【００９２】この発明によれば、文書集合から検索条件に適合する文書を検索し出力する文書検索ステップと、文書に記述されている単語の出現頻度に基づいて、単語の統計的な重みを算出し、重みから得られる文書ベクトル集合を出力する文書特徴抽出ステップと、文書ベクトル集合を、文書ベクトル間の類似度に従って分類することによって話題を作成し、話題に属する文書情報と話題の重要度付き検索語とを出力する話題分類ステップと、文書情報と重要度付き検索語とを参照して話題に属する検索語の絞り込み効果指標を算出し出力する絞り込み効果推定ステップと、絞り込み効果指標を検索語に付与し、絞り込み効果指標が高い検索語を選択して提示検索語候補とし、当該提示検索語候補と当該提示検索語候補に対応する文書情報とを出力する提示検索語候補生成ステップと、提示検索語候補と話題とを対応付けて、絞り込み効果指標と共に提示し、指示情報または選択情報のどちらか一方もしくは両方を入力するように促す分類結果提示ステップと、選択情報に基づいて、文書ベクトル集合に含まれる文書ベクトルを変更して出力する文書特徴設定ステップとを有するように構成したので、絞り込み検索のために提示する提示検索語候補に対して絞り込み効果指標を付与したから、追加検索語の選択を容易にすることができ、絞り込み検索を効率よく実行できるという効果を奏する。

【００９３】この発明によれば、絞り込み効果推定ステップが、少なくとも一つまたは複数の話題に対して絞り込み効果指標を算出し出力するように構成したので、文書単位よりも大きな単位である話題単位で絞り込み検索を実行できるから、より広い範囲の検索結果に対する絞り込み効果指標の傾向が把握できると共に、ユーザが所望する文書の内容を的確に指定できるという効果を奏する。

【００９４】この発明によれば、分類結果提示ステップが、話題と当該話題に属する提示検索語候補とを行列の形式で提示し、行列の各要素に絞り込み効果指標を提示するように構成したので、検索結果の文書全体に対する絞り込み効果の一覧性が高まり、検索結果全体の内容を効率的に把握できるという効果を奏する。

【００９５】この発明によれば、文書検索ステップが出力する文書から当該文書を特徴付ける単語を話題分類項目として抽出し、話題分類項目に関連するテキストを参照して、話題分類項目に対する重みベクトルを学習し、話題分類項目と重みベクトルとを出力する話題分類項目取得ステップを有し、話題分類ステップが、話題分類項目と重みベクトルとに基づいて文書ベクトル集合を分類することによって話題を作成するように構成したので、話題分類ステップで用いる話題分類項目が自動的に取得でき、予め分類カテゴリを設定しておく必要がなくなるから、検索結果の文書項目の設定作業が不要になると共に、絞り込み検索の効率がよくなるという効果を奏する。

【００９６】この発明によれば、話題分類項目取得ステップが、文書検索ステップから出力される文書から当該文書に記述されている単語の出現頻度に基づいて話題分類項目を抽出するように構成したので、話題分類項目を自動的に効率よく抽出することができるという効果を奏する。

【００９７】この発明によれば、話題分類項目取得ステップが、文書検索ステップから出力される文書から当該文書に記述されているタグを参照して話題分類項目を抽出するように構成したので、タグが記述されている文書から話題分類項目を自動的に効率よく抽出することができるという効果を奏する。

【００９８】この発明によれば、分類結果提示ステップを介して指定された文書から文書ベクトルを算出し出力する指定文書特徴抽出ステップを有し、文書特徴設定ステップが、指定文書特徴抽出ステップから出力された文書ベクトルと、文書特徴抽出ステップから出力された文書ベクトル集合とに基づいて文書ベクトル集合を変更するように構成したので、ユーザが所望する文書に近い内容の文書を直接指定することによって、文書特徴設定ステップの文書ベクトルの変更が可能になると共に、絞り込み検索の効率がよくなるという効果を奏する。

【００９９】この発明によれば、所定の単語と関連する単語を定義し関連語として記録する第１の記録ステップと、指定された検索語に対応する関連語を抽出して出力

する関連語設定ステップとを有し、文書特徴設定ステップが、関連語と分類結果提示ステップから入力した選択情報とに基づいて、文書ベクトル集合を変更するように構成したので、検索語の異表記や類似語などを含む文書が検索できるようになるから、検索漏れが抑制されると共に、ユーザが所望する文書を発見するための検索の効率がよくなるという効果を奏する。

【０１００】この発明によれば、検索要求文の作成知識を記録する第２の記録ステップと、提示検索語候補生成ステップが出力した検索語に対応する検索要求文を作成し出力する検索要求作成ステップとを有し、提示検索語候補生成ステップが、絞り込み効果指標に基づいて出力する検索語を選択するように構成したので、初回の検索結果に含まれなかったユーザが所望する文書を文書集合から自動的に検索できるようになると共に、ユーザの絞り込み検索の効率がよくなるという効果を奏する。

【０１０１】この発明によれば、提示検索語候補生成ステップが、複数の検索語を選択して出力し、検索要求作成ステップが、複数の検索語に対する論理演算から検索要求文を作成するように構成したので、ＡＮＤ演算ではより的確に絞り込み検索が実行でき、ＯＲ演算ではより広範囲に絞り込み検索が実行できるという効果を奏する。

【図面の簡単な説明】

【図１】　この発明の実施の形態１による文書検索装置の構成を示すブロック図である。

【図２】　この発明の実施の形態１による文書検索装置の動作を説明するフローチャートである。

【図３】　この発明の実施の形態１における文書ベクトルの一例を示す説明図である。

【図４】　この発明の実施の形態１における検索語−話題対応表の一例を示す説明図である。

【図５】　この発明の実施の形態２による文書検索装置の構成を示すブロック図である。

【図６】　この発明の実施の形態３による文書検索装置の構成を示すブロック図である。

【図７】　この発明の実施の形態４による文書検索装置の構成を示すブロック図である。

【図８】　この発明の実施の形態４における単語と関連語とを定義した一例を示す説明図である。

【図９】　この発明の実施の形態５による文書検索装置の構成を示すブロック図である。

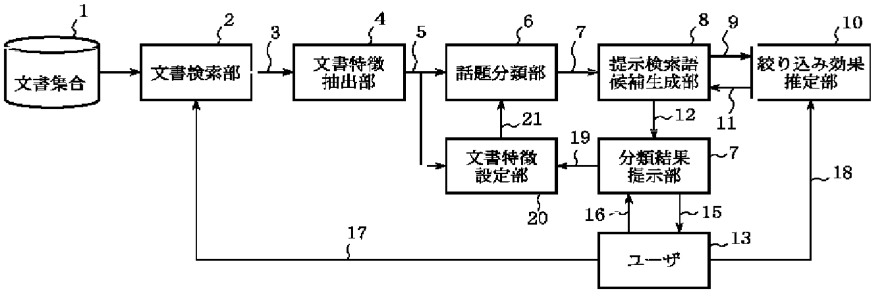
【図１０】　従来の文書検索装置を示す構成図である。

【図１１】　従来の文書検索装置における文献単語行列の一例を示す説明図である。

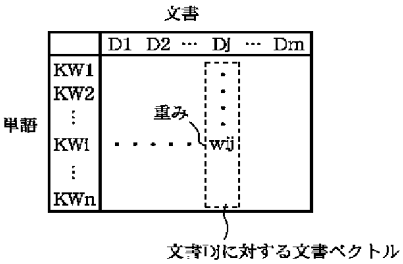
【符号の説明】

１　文書集合、２　文書検索部（文書検索手段）、３　検索結果文書集合、４文書特徴抽出部（文書特徴抽出手段）、５　文書ベクトル集合、６　話題分類部（話題分類手段）、７　重要度付き検索語集合、８　提示検索語候補生成部（提示検索語候補生成手段）、９　重要度付き検索語集合、１０　絞り込み効果推定部（絞り込み効果推定手段）、１１　絞り込み効果指標、１２　提示検索語候補、１３　ユーザ、１４　分類結果提示部（分類結果提示手段）、１５　分類結果、１６　指示情報、１７　検索条件、１８　話題、１９　選択情報、２０　文書特徴設定部（文書特徴設定手段）、２１　文書ベクトル、３１　話題分類項目取得部（話題分類項目取得手段）、３２　話題分類項目情報、４１　指定文書特徴抽出部（指定文書特徴抽出手段）、４２　文書ベクトル、４３　文書、５１関連語辞書（第１の記録手段）、５２　関連語設定部（関連語設定手段）、５３検索語、５４　関連語、６１　検索要求作成知識（第２の記録手段）、６２検索要求作成部（検索要求作成手段）、６３　検索要求検索語、６４　検索要求文。

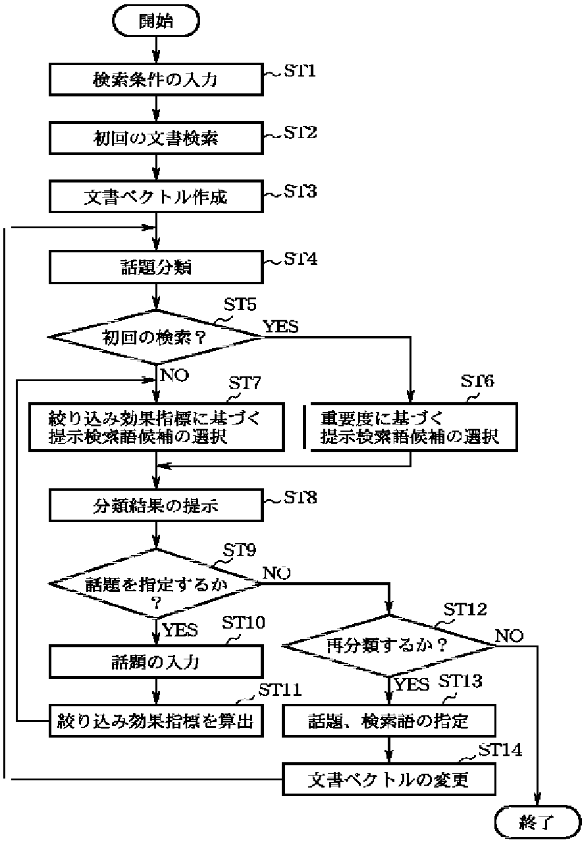
【図１】



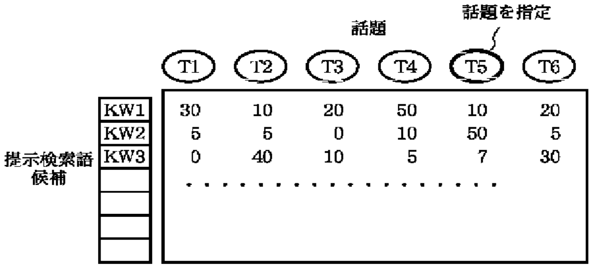
【図３】



【図２】



【図４】

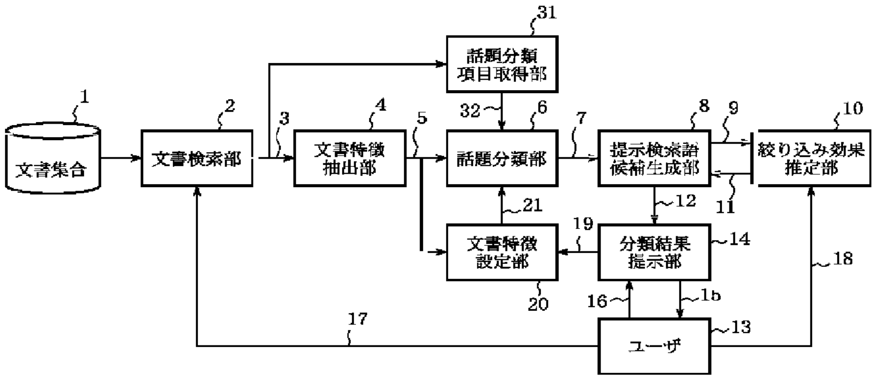


	T1	T2	T3	T4	T5	T6
KW1	30	10	20	50	10	20
KW2	5	5	0	10	50	5
KW3	0	40	10	5	7	30
	・・・・・・・・・・・・・・・・・・					

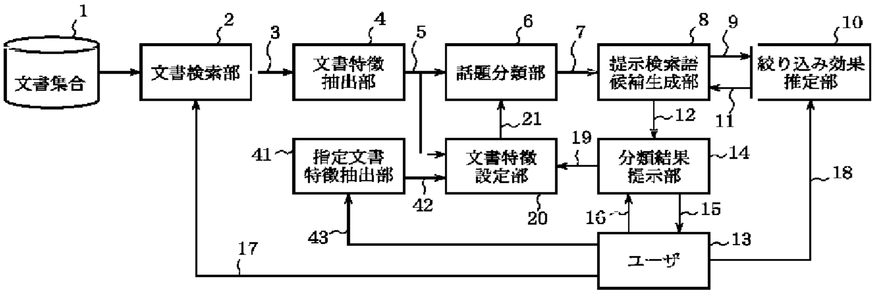
【図１１】

		検索語リスト			
		検索語1	検索語2	検索語3	・・・
文献識別子	文献1	3	0	0	・・・
	文献2	1	2	0	・・・
	文献3	0	0	5	・・・
	⋮	⋮	⋮	⋮	

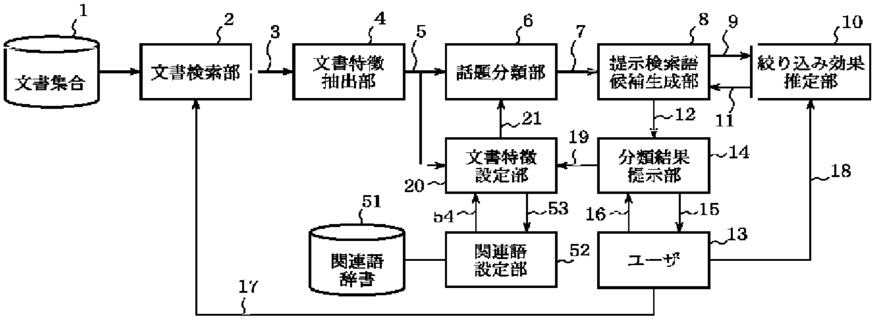
【図５】



【図６】



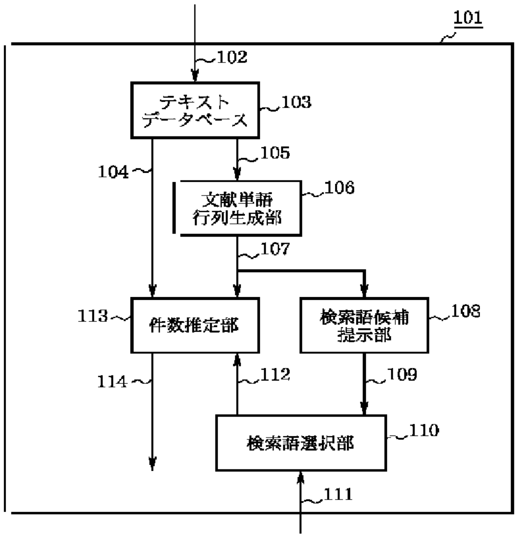
【図７】



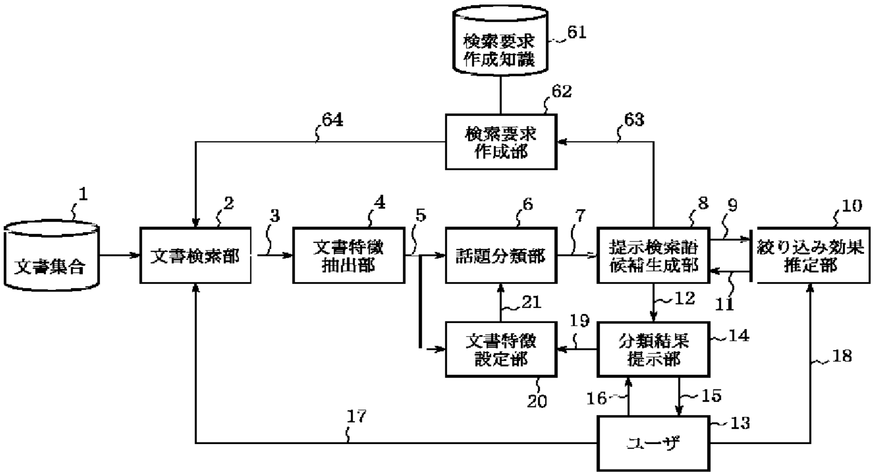
【図８】

単語	関連語	
	異表記	類似語
インターネット	インタネット、inter net	WWW、Web、
ソフトウエア	ソフトウェア、ソフト、S/W	プログラム、アプリケーション
○○電機株式会社	○○電機、○○電機(株)	○○、○○電気、○○電器
本		書籍、ブック、書物、
⋮	⋮	⋮

【図１０】



【図９】



フロントページの続き

(72)発明者　鈴木　克志
東京都千代田区丸の内二丁目２番３号　三
菱電機株式会社内

Ｆターム(参考)　5B075 ND03 NK02 NK31 NR05 NR12
NS01 PP13 PP22 PQ32 PQ46
UU06